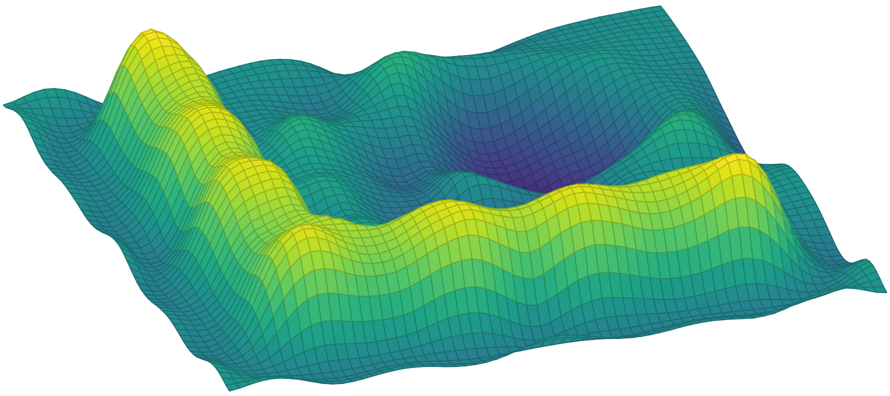
THIES GERKEN

DYNAMIC INVERSE PROBLEMS FOR WAVE PHENOMENA



Dissertation

Dynamic Inverse Problems for Wave Phenomena

submitted by
Thies Gerken

in fulfillment of the requirements
for the degree of *Doctor rerum naturalium (Dr. rer. nat.)*

Center for Industrial Mathematics
University of Bremen
August 2019

Thies Gerken: *Dynamic Inverse Problems for Wave Phenomena*

FIRST ADVISOR:	Prof. Dr. Andreas Rieder, Karlsruhe Institute of Technology
SECOND ADVISOR:	Prof. Dr. Alfred Schmidt, University of Bremen
DATE OF COLLOQUIUM:	October 7, 2019

ABSTRACT

In this work, we deal with second-order hyperbolic partial differential equations that include time- and space-dependent coefficients, and the inverse problems of identifying these coefficients based on their effect on the equation's solution.

We present the needed theory for such equations, including some regularity results for their solution. This allows to state and analyze the inverse problems, even in an abstract setting where time-dependent operators are sought.

Subsequently, we show how these results can be applied to actual partial differential equations. We give a detailed demonstration in the context of the acoustic wave equation. Our results allow the identification of a time- and space-dependent wave speed and mass density in such a setting, and we give an extensive numerical analysis for this case. We also outline how the abstract framework can be applied to other equations, like simple models for electromagnetic waves.

ZUSAMMENFASSUNG

In dieser Arbeit beschäftigen wir uns mit hyperbolischen partiellen Differentialgleichungen zweiter Ordnung, die zeit- und ortsabhängige Koeffizienten beinhalten. Dabei verfolgen wir das Ziel, diese Koeffizienten basierend auf ihrer Wirkung auf die Lösung der Differentialgleichung zu rekonstruieren.

Wir präsentieren die theoretischen Grundlagen, die solche Gleichungen lösbar machen, und beweisen Regularitätsaussagen für ihre Lösungen. Dies erlaubt eine formale Beschreibung des inversen Problems, sogar in einer abstrakten Formulierung, in der zeitabhängige Operatoren die Unbekannten bilden.

Anschließend stellen wir vor, wie sich diese abstrakten Ergebnisse auf partielle Differentialgleichungen anwenden lassen. Dabei legen wir den Fokus auf die akustische Wellengleichung, und der Rekonstruktion von zeit- und ortsabhängiger Wellengeschwindigkeit und Massendichte in dieser Gleichung. Diese inversen Probleme analysieren wir zudem numerisch. Wir zeigen auch auf, welche Ergebnisse für verwandte Gleichungen erzielbar sind, beispielsweise für elektromagnetische Wellen.

PUBLICATIONS

Some of the ideas and results presented in this thesis have also been submitted to scientific journals. We give a brief overview of these publications.

This thesis can be seen as the natural extension of the ideas used in the author's Master's thesis. These results were published as [GL17] (in the special issue on *Dynamic Inverse Problems* [SHB18]) as joint work with Armin Lechleiter. Note that this article was selected as one of 2017's highlight papers of the journal *Inverse Problems*.[1]

The regularity results presented in Section 3.2 are based on joint work with Simon Grützner, and have been accepted for publication in Inverse Problems as [GG19]. They are crucial for the abstract analysis of Chapter 4, which will soon be published as [Ger19], also in Inverse Problems. This article also outlines how the abstract theory can be applied to the elastic wave equation and a simple Maxwell model, which is included as Chapter 6 in this thesis.

The application to the acoustic wave equation, found in Chapter 5, as well as some of the numerical results from Section 8.3, have been submitted to *Springer* as a chapter of an upcoming edited volume on *Time-dependent Problems in Imaging and Parameter Identification*.

[GL17] T. Gerken and A. Lechleiter. "Reconstruction of a Time-Dependent Potential from Wave Measurements." In: *Inverse Problems* 33.9 (2017), p. 094001.

[Ger19] T. Gerken. "Dynamic Inverse Wave Problems – Part II: Operator Identification and Applications." In: *Inverse Problems* (2019).

[GG19] T. Gerken and S. Grützner. "Dynamic Inverse Wave Problems – Part I: Regularity for the Direct Problem." In: *Inverse Problems* (2019).

1 as listed on https://iopscience.iop.org/journal/0266-5611/page/Highlights of 2017 (accessed July 19, 2019)

DANKSAGUNG

Ich möchte Prof. Dr. Armin Lechleiter dafür danken, mich in jeder Abschlussarbeit begleitet zu haben. Er hat mir in Vorlesungen und während meiner Zeit in seiner Arbeitsgruppe gezeigt, wie sich Mathematik nicht nur anwenden, sondern auch ansprechend präsentieren lässt.[2] Das Forschungsthema dieser Dissertation war seine Idee und ich hoffe, er wäre mit dem Ergebnis zufrieden!

Mit seiner Abwesenheit war erstmal nicht klar, wie es mit meiner Promotion weitergeht. Zu meinem Glück hat sich Prof. Dr. Andreas Rieder bereiterklärt, die Erstbetreuung aus der Ferne zu übernehmen. Er hat regelmäßig einen vollständigen Arbeitstag mir und meiner Promotion gewidmet. Das wäre auch in einem regulären Betreuungsverhältnis nicht selbstverständlich! Prof. Dr. Alfred Schmidt hat sich dankenswerter Weise bereiterklärt, seine Rolle als mein Zweitbetreuer und Gutachter über die Masterarbeit hinaus fortzusetzen. So hatte auch vor Ort jemand ein Auge auf meinen Fortschritt.

Durch meine Zugehörigkeit zum DFG-finanzierten Graduiertenkolleg 2224 (π^3: Parameter Identification — Analysis, Algorithms, Implementations) hatte ich auch nach der schleichenden Auflösung der Arbeitsgruppe weiterhin einen festen Platz im Zentrum für Technomathematik. Danke an meine Kollegen dort für die interessanten Kaffee- und Mittagspausen!

Nicht zuletzt möchte ich meiner Familie dafür danken, mir das Studium ermöglicht zu haben, und meiner Freundin, da sie mich ermutigt hat auch noch die Promotion anzuhängen.

2 Auch wenn ich weiterhin überzeugt bin, dass Summen in Integranden keiner Klammerung bedürfen ;-)

CONTENTS

1 INTRODUCTION 1
2 LINEAR EVOLUTION EQUATIONS OF SECOND ORDER 5
2.1 Preliminaries . . . 7
2.2 Existence of solutions . . . 11
2.3 Uniqueness of solutions . . . 20
3 REGULARITY RESULTS 24
3.1 Lions' Lemma . . . 25
3.2 Differentiating the equation with respect to time . . . 32
3.3 Conclusions . . . 41
4 ABSTRACT OPERATOR FRAMEWORK 44
4.1 Well-posedness of the forward operator . . . 45
4.2 Fréchet-differentiability . . . 49
4.2.1 Validity of the tangential cone condition . . . 54
4.2.2 Total differentiability . . . 55
4.2.3 Adjoint of the derivative . . . 58
4.3 Ill-posedness . . . 61
5 APPLICATION TO THE ACOUSTIC WAVE EQUATION 68
5.1 Pursuit of a suitable equation . . . 69
5.1.1 The wave speed . . . 70
5.2 Construction of the forward operator . . . 75
5.2.1 Injectivity . . . 79
5.3 Fréchet-differentiability . . . 79
5.3.1 Adjoint of the derivative . . . 84
5.4 Ill-posedness . . . 87
6 APPLICATION TO OTHER EQUATIONS 93
6.1 Linear elasticity . . . 93
6.1.1 Differentiability . . . 97
6.1.2 Ill-posedness . . . 98
6.2 Maxwell's equations . . . 100
6.2.1 Differentiability . . . 102
6.2.2 Ill-posedness . . . 103
7 DISCRETIZATION OF THE ACOUSTIC WAVE EQUATION 105
7.1 Solver for the partial differential equation . . . 106
7.1.1 Semidiscretization in time . . . 108
7.1.2 Discretization in space . . . 112
7.1.3 Numerical validation . . . 116
7.2 Inner products . . . 119
7.2.1 Approximation of $L^2(I; L^2(\Omega))$ . . . 119
7.2.2 Including time derivatives . . . 121
7.3 Discretized operators . . . 122
8 NUMERICAL INVERSION OF THE ACOUSTIC WAVE EQUATION 129
8.1 Regularization approach . . . 129
8.1.1 Reformulating the problem in Hilbert spaces . . . 131

8.1.2 Regularization methods 132
8.1.3 Admissibility of iterates 136
8.2 Inversion scenarios 138
8.3 Convergence 141
8.3.1 Smooth parameter in $H^1(I; L^2(\Omega))$ 143
8.3.2 Influence of the reconstruction norm 146
8.3.3 Violating the smoothness constraints 148
8.4 Reconstruction from partial data 150
8.4.1 Static reconstructions 157
9 CONCLUSION & OUTLOOK 159
A BOCHNER SPACES 161
B NUMERICAL CONVERGENCE RATES 167

BIBLIOGRAPHY 169

LIST OF FIGURES

Figure 4.1 Decomposition of a possible forward operator F . 44
Figure 5.1 Initial values u_0 as defined in (5.12) 72
Figure 5.2 Observed wave speeds of u_{in}, u_{out} and u, together with the parameter c as given in (5.13) 73
Figure 5.3 Approximations of u_{in}, u_{out} and u as they leave $(-11, 11)$ 74
Figure 7.1 Structuring of space-time meshes of line methods 107
Figure 7.2 Convergence of errors in the 2D reference scenario with piecewise linear elements 118
Figure 7.3 The approximation of the adjoint and the adjoint of the approximation 123
Figure 8.1 Inverse transformation γ^{-1} 138
Figure 8.2 Test parameter Λ_{LDot}, evaluated at $t = \pi/2$ 139
Figure 8.3 Test parameter Λ_{Ring}, evaluated at $t = \pi$ 140
Figure 8.4 Data of the initial guess compared to the exact data for $c = 0.15\Lambda_{\text{LDot}}$ 141
Figure 8.5 Reconstruction of $c = 0.15\Lambda_{\text{LDot}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-2}$ 143
Figure 8.6 Reconstruction of $c = 0.15\Lambda_{\text{LDot}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$ 143
Figure 8.7 Errors and computing times for reconstruction of 2D $c = 0.15\Lambda_{\text{LDot}}$ using $H^1(I; L^2(\Omega))$ depending on the noise level 144
Figure 8.8 Reconstruction of $\rho = 0.75\Lambda_{\text{LDot}}$ in $H^1(I; L^2(\Omega))$ in 3D with $\varepsilon = 4.6 \cdot 10^{-2}$ 145
Figure 8.9 Reconstruction of $\rho = 0.75\Lambda_{\text{LDot}}$ in $H^1(I; L^2(\Omega))$ in 3D with $\varepsilon = 10^{-3}$ 145
Figure 8.10 Reconstruction of $q = 20\Lambda_{\text{LDot}}$ in $H^1(I; L^2(\Omega))$ in 3D with $\varepsilon = 10^{-3}$ 146
Figure 8.11 Errors for 2D $q = 20\Lambda_{\text{LDot}}$ for different reconstruction norms 147
Figure 8.12 Discrepancies of REGINN-iterates during reconstruction of Λ_{LDot} using $L^2(I; L^2(\Omega))$ and $\varepsilon = 10^{-4}$ 148
Figure 8.13 Errors for 2D $\nu = 10\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$ 149
Figure 8.14 Reconstruction of $\nu = 10\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$ 149
Figure 8.15 Reconstruction of $\nu = 10\Lambda_{\text{Ring}}$ in $L^2(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$ 149
Figure 8.16 Last iterate for $c = 0.15\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$ 150
Figure 8.17 Reconstruction of $c = 0.15\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-2}$ 150

Figure 8.18 Data $[\mathbf{F}_c(0.15\Lambda_{\mathrm{LDot}})]_2$ for different measurement operators Ψ . . . 154
Figure 8.19 Reconstruction of $c = 0.15\Lambda_{\mathrm{LDot}}$ using different measurement operators Ψ and 0.1% noise . . . 155
Figure 8.20 Relative $L^2(\Omega)$-reconstruction error over time for $c = 0.15\Lambda_{\mathrm{LDot}}$ and 0.1% noise . . . 156
Figure 8.21 Reconstruction of $\rho = 0.75\Lambda_{\mathrm{LDot}}$ from internal measurements and 0.1% noise . . . 157
Figure 8.22 Reconstruction of $\rho = 0.75\Lambda_{\mathrm{LDot}}$ from boundary measurements and 0.1% noise . . . 157
Figure 8.23 Reconstructions of static parameters in $L^2(\Omega)$ in 2D with $\varepsilon = 10^{-3}$ from boundary measurements . 158

LIST OF TABLES

Table 7.1 Values of Δ_{adj} in 2D with two different sets of parameters . . . 127
Table B.1 Rates for Λ_{LDot} in 2D . . . 167
Table B.2 Rates for Λ_{Ring} in 2D . . . 168
Table B.3 Rates for Λ_{LDot} using $H^1(I; L^2(\Omega))$ in 3D . . . 168
Table B.4 Rates for Λ_{Ring} using $H^1(I; L^2(\Omega))$ in 3D . . . 168

1 INTRODUCTION

Let us imagine a box filled with water, and a small quantity of other chemicals concentrated somewhere in this container that is moving through it. The only way to see inside is via some speakers and microphones near the container's boundary, which can be used to excite and measure acoustic waves in the water. We know that the chemicals change the physical parameters of the water, for example they might have a different mass density than water or alter the speed at which the sound waves propagate. Is this setup enough to track the distribution of the chemicals over time?

The difficulty in this problem is the fact that the sound waves only indirectly determine the searched for density or wave speed. In other words, a causal connection is more naturally found in the reverse direction: a change in these parameters directly translates to changes in the wave field near the microphones. This relation is much easier to analyze on a theoretical level and to simulate numerically. Mathematically, this yields a *forward operator* $F\colon X \to Y$, that maps parameters $x \in X$ to their "effects" $y \in Y$. The original task consists of solving the *inverse problem* of finding $x \in X$ that solves $Fx = y$ for given data $y \in Y$. Unfortunately, it is common for such problems to be *ill-posed* in the sense that (i) no such x exists, (ii) multiple solutions x exist, or (iii) there is a unique x, but it does not continuously depend on the data y. Especially the last condition can gravely complicate the practical inversion of F since the exact data is hard to come by. It is usually tainted by small measurement errors and numerical inaccuracies, which are amplified by the ill-posedness of F and thereby lead to worthless reconstructions.

The introductory example can be modeled as an inverse problem in which the forward operator F is connected to the solution of a hyperbolic partial differential equation. In this case, this is the acoustic wave equation

$$\frac{1}{\rho(t,x)}\frac{d}{dt}\left(\frac{u'(t,x)}{c(t,x)^2}\right) - \operatorname{div}\left(\frac{\nabla u(t,x)}{\rho(t,x)}\right) = f, \tag{1.1}$$

complemented with suitable initial- and boundary conditions. The unknown time- and space-dependent parameters are the wave speed c and mass density ρ, and they determine the coefficients of this PDE. The data is the wave field u, observed on a bounded time interval. It is generated by exciting a wave using a known volumetric force f. This yields the forward operator $F\colon (c,\rho) \mapsto u$.

Since the searched for quantities in this scenario are time-dependent, we can also refer to it as a *dynamic inverse problem*. The additional dimension in the unknowns complicates the theoretical analysis of the direct problem. Furthermore, it increases the computational resources (both processing time and memory storage) that suitable inversion algorithms consume. The corresponding problems for static parameters are not easy to solve either (cf. [KR14b]). Thus, most of the literature deals with these conventional inverse problems, while the treatment of dynamic inverse problems is a relatively new area of research. In this work, we focus on a subclass of time-dependent problems, namely those that are governed by some kind of wave propagation.

Apart from the fact that we only deal with *linear* PDEs, the example of moving fluids fulfills another restriction that we continue to make throughout this work. Through the modeling using a single PDE we implicitly assumed that the speed in which the parameters change approximately matches the speed of the wave propagation. Were the unknowns to vary much more rapidly, then we could not expect that the wave field is able to resolve these changes. In the other extreme of a very slowly-changing parameter, the wave fronts simply rush by and reach the sensors without being influenced by any dynamic behavior. In the latter case one might as well use a conventional reconstruction algorithm that assumes the unknowns to be constant in time.

APPLICATIONS Naturally, the assumptions above restrict the set of possible applications. For example computerized tomography of the human body does not belong to this class, since all imaginable motion (breathing, heartbeats) is magnitudes slower than that of the X-rays, which travel at light speed. The movement becomes significant only due to the pauses between measurement events, in which the device is repositioned. Still, there are some scenarios that fit into our framework. This is even the case for tomography, as long as the object of interest moves very fast, which could be the case for plasma. Moreover, recent advances in photonic structures have made it possible to create structures that rapidly change their electric permittivity when an external field is applied (cf. [Hay+16] and the references therein). This change in permittivity can be measured using electromagnetic waves.

Let us substitute the water-filled container with a solid material, and attach piezo-electric crystals on its surface. These can be used to excite and measure elastic waves inside this material. Inhomogeneities will scatter or reflect these waves, thus this setup can be used in nondestructive testing (NDT). There, this solid object is likely to be some kind of carbon fiber reinforced polymer which has to be inspected for structural damage (cf. [LS17]). These kinds of materials are frequently used in aviation. A significant time-dependence comes into play if this measurement takes place during a tension test, with the goal of observing not the damage itself, but the *process* by which it is formed (see [GM18]). Other possible applications include situations where the probing wave interacts with the material under test, for example by heating it, which changes the mass density.

Note that the previously mentioned areas of applications involve waves, but the acoustic wave equation (1.1) is not the correct tool to model them. For example, deformations of solid objects (as encountered in nondestructive testing) are better described by the elastic wave equation, and electromagnetic waves behave according to Maxwell's equations.

OUR APPROACH The starting point of our research was the identification of the potential q in the wave equation $u'' - \Delta u + qu = f$. This problem is much easier to analyze, because it does not influence terms that involve derivatives (in space or time) of u. The same reasoning applies to the numerical computation of u and the solution of the inverse problem. Indeed, the analysis even works for parameters q that are merely square-integrable functions of space and time. Furthermore, the corresponding numerical results appeared to be very promising (see [Ger16; GL17]).

The natural next step was to take up the acoustic wave equation (1.1), and only then move on to the other, above mentioned, equations. After focusing on the acoustic wave equation, we discovered that our approaches to this problem were not really specific to this equation at all. Instead, we "accidentally" based our analysis largely on its representation as an *evolution equation*

$$\frac{d}{dt}\big(C(t)u'(t)\big) + B(t)u'(t) + A(t)u(t) + Q(t)u(t) = f(t). \tag{1.2}$$

In particular, we saw that the treatment of other equations could be performed by building upon the same underlying ideas. Hence, we then pursued the idea of extracting these ideas in a way that they could *directly* be applied to other problems. This gave rise to a general framework that separates the problem-dependent behavior from the treatment of the abstract evolution equation (1.2). We did this by regarding the linear operators A, B, C and Q in the evolution equation as the arguments to a new "abstract" inverse problem with forward operator $S\colon (A, B, C, Q) \mapsto u$. We performed a rigorous analysis of this nonlinear operator. This includes the setup of suitable function spaces and well-posedness results for the wave field u. Furthermore, we could show that this operator is Fréchet-differentiable, and even characterized the corresponding adjoints in the general setting. As suspected, even this abstract problem is ill-posed.

How the operators A, B, C and Q have to be set up depends on the problem at hand. With little effort one can transfer the properties of S to specific problems, provided that they can be written in the form of (1.2). Consequently, we also reformulated our results for the acoustic wave equation so that they build on this framework. To provide a proof of concept, we also used it for dynamic problems concerning the elastic wave equation and a model for electromagnetic waves.

In addition to the theoretical treatment, we developed a software library for the numerical solution of dynamic inverse problems that involve the acoustic wave equation. The title page of this thesis shows a snapshot of a reconstructed time-dependent mass density that was obtained with our software.

CONNECTIONS TO EXISTING WORK The corresponding tasks with *time-independent* parameters have already been discussed in detail, for example by [KR14a; KR14b; LS17], and there exist general frameworks for their treatment [BSS13; KR16]. Our results give the means to pursue the related dynamic inverse problems.

There are some articles concerning uniqueness for the identification of a time-dependent potential [Aïc15; HK18; Kia17; KO17; RS91; RR91; Ste89; Wat14]. However, to our knowledge there are no articles that deal with time-dependent coefficients close to the second-order derivatives. In particular, this means that our rigorous analysis of the direct problems, and the proof of their Fréchet-differentiability, is state of the art. Furthermore, the attempt of a numerical reconstruction of such (possibly four-dimensional) parameters in an inverse problems context is a novelty.

Recent advances in dynamic inverse problems mostly deal with time-dependent behavior that is much slower than the speed of wave propagation [Hah13; Hah14; Hah17], and they often include prior-knowledge about the time-dependence. We do not restrict the structure of the time-dependence in the parameters. In our setup, the time frames of the unknowns are only loosely coupled through the involved function spaces.

STRUCTURE OF THE THESIS We start by getting comfortable with the theory for the evolution equation (1.2). Furthermore, we extend the known existence and uniqueness results for its solution to the time-dependent case. The treatment of the inverse problems hinges on regularity results. Most usual approaches cannot be used in case of time-dependent operators, thus we devote Chapter 3 to the quest for suitable regularity theorems. Chapter 4 deals with the aforementioned abstract inverse problems, and it heavily builds upon the well-posedness- and regularity results from the preceding chapters. With it, we can tackle specific partial differential equations. Our main focus is the acoustic wave equation (1.1). The task of finding the wave speed, the mass density and two additional parameters is discussed in Chapter 5. In Chapter 6 we demonstrate that analogous problems with other PDEs can also be solved using our framework. We do so by applying it to the elastic wave equation and a model based on Maxwell's equations. In Chapter 7 we return to the acoustic wave equation and discuss its numerical approximation. This includes numerical schemes to approximate the adjoint of the forward operator's Fréchet-derivative. This is the preparation for Chapter 8, where an inexact Newton method is used to numerically solve example problems. Among other things, we discuss whether these reconstructions converge if the data quality increases. We close this thesis by giving some suggestions for future research.

2 LINEAR EVOLUTION EQUATIONS OF SECOND ORDER

In this chapter we discuss theory for evolution equations that have the form

$$\frac{d}{dt}\big(C(t)u'(t)\big) + B(t)u'(t) + A(t)u(t) + Q(t)u(t) = f(t) \tag{2.1}$$

in order to be able to rely on these results in subsequent chapters.

The evolution equation above is "abstract" in the sense that we view the underlying partial differential equation as a kind of *ordinary* differential equation. This is accomplished by hiding the spatial derivatives in the fact that the operators in the equation map between different Hilbert spaces V and H. We assume that $V \subset H$ (in the sense that the embedding $V \hookrightarrow H$ is continuous) and V is dense in H. Therefore we can see elements belonging to V as more regular than those that belong to H. In applications to specific PDEs (e.g. in Chapters 5 and 6 of this work) H will be some kind of L^2 space, while functions in V additionally have well-defined spatial derivatives. Moreover, the space V may enforce boundary conditions.

We identify the dual space H^* of H with the space itself, therefore $H \subset V^*$ by $u \mapsto (u, \cdot)$. Since V is reflexive, H is a dense subset of V^*. We obtain $V \subset H \subset V^*$, where both embeddings are dense. This kind of relationship between V, H and V^* is commonly referred to as them forming a *Gelfand-* or *evolution triple*. Evolution triples play an important role in the treatment of parabolic and hyperbolic equations. For example, they allow integration by parts for time-dependent functions that have values in V, but are only differentiable with respect to H. Note that although we require V to be a Hilbert space, we will not make direct use of its inner product. However, we sometimes need an integration by parts formula for functions that take values in H and are differentiable in V^*. Because V is a Hilbert space, $H \subset V^* \subset H^*$ forms another evolution triple. Thus, the integration by parts formula is also valid in this case.

We further require that V and H are separable spaces. The main reason for this is the fact that the Galerkin method, which we use to show existence of u, does not work without the existence of orthonormal bases.

We would like to note that it is a bad idea to use the inner product of V to identify V^* with V, because this notation would then tempt us to conclude $H \subset V$. The error in this argument is not immediately apparent. It is hidden in the fact that "$\subset$" gets assigned two meanings that are based on different embeddings. By writing $H \hookrightarrow V^* \hookrightarrow V$ instead, we see that we have merely constructed a map $i\colon H \to V$ and confused H and $i(H)$, although $i(H)$ and H are not isometrically isomorphic.[1]

1 A closer look reveals that i is the adjoint to the embedding $V \hookrightarrow H$.

Throughout the remainder of this chapter we assume that V and H are separable Hilbert spaces that yield a Gelfand triple $V \subset H \subset V^*$. Because the embedding $V \hookrightarrow H$ is continuous we can (without loss of generality) assume that $\|u\|_H \leqslant \|u\|_V$ holds for all $u \in V$.

In subsequent chapters we will encounter inverse problems where u represents the data. It is unrealistic that measurements of $u(t)$ are available on an unbounded time domain. Furthermore, unless the time-dependence of the operators and the right-hand side in (2.1) is simple (for example time-harmonic), $u(t)$ will neither be periodic in t, nor will it be approximately equal to some periodic function for large times t. Therefore it is safe to assume that in possible real-world applications, $u(t)$ will not be known on an unbounded time domain; more realistically, measurements of u at a finite number of time instances will be available. Hence we will complement (2.1) with initial conditions and solve it on a bounded time interval. For convenience we assume that this interval is given as $I = [0, T]$ with a fixed positive end time T.

We use the first section of this chapter to get acquainted with equation (2.1) by considering suitable initial conditions and requirements that ensure a well-posed problem statement. We continue in Section 2.2 by discussing necessary assumptions on the data and the operators A, B, C and Q that guarantee solvability of the resulting initial value problem. In Section 2.3 we then prove that the solution we have constructed must be the only one.

For all of our analyses we rely on energy methods. Evolution equations can equivalently be analyzed using the semigroup framework, see for example [Paz83; Sho97]. However, we feel that especially equations full of time-dependent operators are easier to analyze using energy methods.

Furthermore, apart from the numerical examples, we will always use the second order formulation (2.1). One could equivalently look at the corresponding system of first-order equations, but to the author's knowledge this has little advantage.

NOTATION We wish to make a few remarks on our notation. We will make heavy use of the usual Bochner space of time-dependent functions $u \colon I \to X$ that take values in a Banach space X throughout this thesis. We denote these kinds of spaces via $W^{k,p}(I; X)$ and $L^p(I; X)$. For the convenience of the reader we have collected their definitions as well as some important properties in Appendix A.

By $(\cdot, \cdot)$ and $\langle \cdot, \cdot \rangle$ we denote the inner product of H and the dual product of V^* and V, respectively. The use of other inner- or dual products will be indicated by a subscript or explicitly stated in the text. Further, we write $\mathcal{L}(X, Y)$ for the space of linear and continuous operators between normed spaces X and Y, with the shorthand notation $\mathcal{L}(X)$ if $X = Y$.

For operators belonging to $L^\infty(I; \mathcal{L}(X, Y))$ we denote their *realization* using calligraphic font. To be more precise, for $A \in L^\infty(I; \mathcal{L}(X, Y))$ the operator $\mathcal{A} \colon L^2(I; X) \to L^2(I; Y)$ is defined by

$$(\mathcal{A}v)(t) := A(t)v(t),$$

which is valid for almost all $t \in I$ if $v \in L^2(I;X)$. This notation is very handy when trying to write (2.1) not pointwise in time, but as an equality of time-dependent functions. For instance, it enables us to write $\mathcal{A}v = 0$ as an equality in $L^2(I;Y)$, as opposed to stating $A(t)v(t) = 0$ in Y for almost all $t \in I$. With $g^{(k)}$ we usually denote the kth classical, weak or distributional derivative of the function g, depending on the context. For lower order derivatives we also write g' or g''. In the case of multiple variables, this refers to the derivative with respect to time. The main exception to these rules is $\mathcal{A}^{(k)}$, which we define to be the realization of $A^{(k)}$, i.e. the kth derivative of the time-dependent operator A. This should not cause any confusion, because the aforementioned types of derivatives cannot be defined for $\mathcal{A}$. Another useful shorthand notation, which we will sometimes employ is $k_1 := \max\{1, k\}$ for $k \in \mathbb{N} \cup \{0\}$. Apart from the fact that we denote the set of all compactly supported smooth functions as $C_c^\infty(I)$ (and $C_c^\infty(I;X)$), and the support of a function f by $\operatorname{spt} f$, all our remaining notation should be fairly standard or at least self-explanatory.

2.1 PRELIMINARIES

Let us first fix the idea of a Gelfand triple (as motivated in the introduction) in a definition.

Definition 2.1. Let V, H be separable Hilbert spaces, where V is continuously and densely embedded in H. By identifying H with H^* we obtain the relationship

$$V \subset H \subset V^*,$$

which we denote as a *Gelfand triple*.

We have not yet specified in what way u should solve (2.1). We expect the solution to be an element of $L^2(I;V)$, with the first time derivative u' belonging to $L^2(I;H)$, i.e. $u \in L^2(I;V) \cap H^1(I;H)$. The structure of the operators is as follows: B, $C \in L^\infty(I;\mathcal{L}(H))$, while $Q \in L^\infty(I;\mathcal{L}(V,H))$ and $A \in L^\infty(I;\mathcal{L}(V,V^*))$. In applications to actual PDEs this means that $A(t)$ may encode second-order spatial derivatives (and in fact must do so in order to be coercive), whereas $Q(t)$ can still incorporate first-order spatial derivatives. These conditions are not "minimal", they already take into account what kind of problems we aim to solve in subsequent chapters. For example, it is possible to obtain L^2-theory for Q, at least in the special case of the wave equation (cf. [GL17]), and $B(t) \in \mathcal{L}(V,H)$ is sufficient after modifying the equation to use B^* instead of B (cf. [Lio61]).[2] The right-hand side f should at least belong to $L^2(I;V^*)$. With these assumptions (2.1) is well-defined as an equality in V^*, and we can also write it as

$$(\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{A} + \mathcal{Q})u = f \quad \text{in } L^2(I;V^*). \tag{2.2}$$

2 In the context of the wave equation this would allow us to have $\operatorname{div}(\beta u')$ in the equation.

By regarding $\mathcal{C}u' \in L^2(I;V^*)$ as a distribution $C_c^\infty(I) \to V^*$, its derivative $(\mathcal{C}u')' \in L^2(I;V^*)$ is always well-defined in the distributional sense via $(\mathcal{C}u')'[\varphi] = -\int_0^T (Cu')(t)\varphi'(t)\,dt \in V^*$ for all $\varphi \in C_c^\infty(I)$. Then the equation above can be understood in the sense that this derivative should be equal to the regular distribution that is associated to $f - \mathcal{B}u' - (\mathcal{A} + \mathcal{Q})u \in L^2(I;V^*)$. It is of no consequence whether we do this or see $\mathcal{C}u' \in H^1(I;V^*)$ as an additional constraint that has to be fulfilled by u.

The only information about second time derivatives of u that we will be able to obtain will stem from the equation itself, i.e. even with $C = \mathrm{Id}$ we would only be able to conclude $u \in H^2(S;V^*)$. This is the reason why we chose the formulation (2.2) with leading term $(\mathcal{C}u')'$. Had we used $\mathcal{C}u''$ instead, then the appropriate space for $C(t)$ would have been $\mathcal{L}(V^*)$. Working with this space, and defining suitable operators in it is awkward at best, and is certainly more difficult than dealing with $\mathcal{L}(H)$. However, we can get results for this kind of equations (with $C(t) \in \mathcal{L}(H)$) by using regularity results, as will be demonstrated in Section 3.3.

We also need to discuss suitable initial conditions for (2.2). Normally, we would require $u(0) = u_0$ and $u'(0) = u_1$, but due to the presence of C the evaluation of u' at $t = 0$ might not be well-defined. For Banach spaces X, we know of the embedding $H^1(I;X) \hookrightarrow C(I;X)$ (see Theorem A.10) and use this to always identify equivalence classes in $H^1(I;X)$ with their continuous members. Together with $u \in H^1(I;H)$ this lets us conclude that $u(0)$ is well-defined as an element of H. We have no knowledge about u'' and therefore cannot give a similar argument for $u'(0)$. Since $\mathcal{C}u' \in H^1(I;V^*)$ we can however use $(\mathcal{C}u')(0) = u_1 \in V^*$ as a second initial condition which is well-defined as an equality in V^*. We would like to remark that $(\mathcal{C}u')(0)$ is (in general) not the same as $C(0)u'(0)$, as the latter would again imply that u' could be evaluated at zero.

Note that these are only preliminary conditions on the operators, the right-hand side and even the initial values. They only allow the statement of the problem; we do not claim (nor expect) that these conditions already ensure the existence of u.

The fact that we expect (2.2) to hold for almost all $t \in I$ in the V^* sense already suggests that we are dealing with a variational problem. The following lemma shows that the corresponding problem with bilinear forms is equivalent to the solution of (2.2); the choice which notation to use is merely a matter of personal preference. Suitable bilinear forms can be defined for almost all $t \in I$, $\varphi, \psi \in V$ and $v, w \in H$ via

$$
\begin{aligned}
&a\colon I \times V \times V \to \mathbb{R}, && a(t,\varphi,\psi) := \langle A(t)\varphi, \psi\rangle + (Q(t)\varphi, \psi),\\
&b\colon I \times H \times H \to \mathbb{R}, && b(t,v,w) := (B(t)v, w),\\
&c\colon I \times H \times H \to \mathbb{R}, && c(t,v,w) := (C(t)v, w).
\end{aligned}
$$

Additional conditions that have to be imposed in the sequel can directly be translated between the two formulations, for instance symmetry of $c(t,\cdot,\cdot)$ is equivalent to C being pointwise self-adjoint.

Lemma 2.2. *The function* $u \in H^1(I;H) \cap L^2(I;V)$ *solves* (2.2) *if and only if it solves*

$$\frac{d}{dt}c(t,u'(t),\varphi) + b(t,u'(t),\varphi) + a(t,u(t),\varphi) = \langle f(t),\varphi\rangle \tag{2.3}$$

for all $\varphi \in V$ *in the weak sense for almost all* $t \in I$.

We would like to point out that it is very important that (2.3) holds for all φ and almost all $t \in I$, and not the other way around. Otherwise the subset of measure zero of I, in which the formulation does *not* hold, would be allowed to depend on φ. In particular, it could happen (assuming $V \neq \{0\}$) that there exists not even a single $t \in I$, for which the formulation holds for every $\varphi \in V$.

Proof. The main, but subtle difference between the two formulations is the following: At least on the first glance, formulation (2.3) only asserts the validity of the differential equation for every fixed test function φ. In contrast, equation (2.2) requires it to hold uniformly for all test functions. First, we note that the operator formulation is (by the definition of distributional derivatives) equivalent to

$$\begin{aligned} -\int_0^T & C(t)u'(t)\psi'(t)\,dt \\ &= \int_0^T \big[f(t) - B(t)u'(t) - A(t)u(t) - Q(t)u(t)\big]\psi(t)\,dt \end{aligned} \tag{2.4}$$

being satisfied for all $\psi \in C_c^\infty(I)$. The integral on either side of this equation is a Bochner integral and yields a value in V^*. Equality in V^* is just pointwise equality at all $\varphi \in V$. We test (2.4) with such a φ and move the dual product into the integral (cf. Theorem A.5) to end up with

$$\begin{aligned} -\int_0^T & c(t,u'(t),\varphi)\psi'(t)\,dt \\ &= \int_0^T \big[\langle f(t),\varphi\rangle - b(t,u'(t),\varphi) - a(t,u(t),\varphi)\big]\psi(t)\,dt, \end{aligned}$$

which is the weak formulation of (2.3). □

An important ingredient used throughout this chapter is the following "product rule" for time-dependent linear operators.

Lemma 2.3. *Let* X, Y *be separable Banach spaces and* $G \in W^{1,\infty}(I;\mathcal{L}(X,Y^*))$. *Given any* $u \in H^1(I;X)$, $v \in H^1(I;Y)$, *the map* $t \mapsto \langle G(t)u(t), v(t)\rangle$ *belongs to* $W^{1,1}(I)$ *and we have*

$$\begin{aligned} \frac{d}{dt}&\langle G(t)u(t),v(t)\rangle \\ &= \big\langle G'(t)u(t),v(t)\big\rangle + \big\langle G(t)u'(t),v(t)\big\rangle + \big\langle G(t)u(t),v'(t)\big\rangle \end{aligned} \tag{2.5}$$

for almost all $t \in I$. *Here,* $\langle\cdot,\cdot\rangle$ *denotes the dual product between* Y^* *and* Y.

Proof. It is clear that both the right-hand side of (2.5) and the mapping $t \mapsto \langle G(t)u(t), v(t)\rangle$ belong to $L^1(I)$.

Since G is weakly differentiable, we know that it also has to be continuous (cf. Theorem A.10). Furthermore, its continuous representative fulfills $(G(t+h) - G(t))/h = \int_t^{t+h} G'(s)\,ds/h \in \mathcal{L}(X, Y^*)$ for all $t \in (0,T)$ and small enough $h > 0$. The right-hand side of this equation converges to $G'(t)$ (due to Lemma A.7), at least for almost all $t \in (0,T)$, when $h \to 0$. Hence, G is almost everywhere differentiable in the *classical* sense. Due to Lemma A.9, there exist $u^\varepsilon \in C^\infty(I;X)$, $v^\varepsilon \in C^\infty(I;Y)$ such that $u^\varepsilon \to u$ in $H^1(I;X)$ and $v^\varepsilon \to v$ in $H^1(I;Y)$ as $\varepsilon \to 0$. We can apply the chain rule to infer

$$\begin{aligned}\frac{d}{dt}&\langle G(t)u^\varepsilon(t), v^\varepsilon(t)\rangle \\ &= \big\langle G'(t)u^\varepsilon(t), v^\varepsilon(t)\big\rangle + \big\langle G(t)(u^\varepsilon)'(t), v^\varepsilon(t)\big\rangle + \big\langle G(t)u^\varepsilon(t), (v^\varepsilon)'(t)\big\rangle\end{aligned}$$

in the classical sense. In particular, we are able to conclude that

$$\begin{aligned}-\int_0^T &\langle G(t)u^\varepsilon(t), v^\varepsilon(t)\rangle \psi'(t)\,dt \\ &= \int_0^T \big\langle G'(t)u^\varepsilon(t), v^\varepsilon(t)\big\rangle \psi(t)\,dt \\ &\qquad + \int_0^T \big[\big\langle G(t)(u^\varepsilon)'(t), v^\varepsilon(t)\big\rangle + \big\langle G(t)u^\varepsilon(t), (v^\varepsilon)'(t)\big\rangle\big]\psi(t)\,dt\end{aligned}$$

holds for all $\psi \in C_c^\infty(I)$. Both sides of this equation converge to the respective terms evaluated at u and v as $\varepsilon \to 0$. We demonstrate this on the first expression on the right-hand side by the straightforward computation

$$\begin{aligned}&\left|\int_0^T \big[\big\langle G'(t)u(t), v(t)\big\rangle - \big\langle G'(t)u^\varepsilon(t), v^\varepsilon(t)\big\rangle\big]\psi(t)\,dt\right| \\ &\leqslant \int_0^T \big|\big[\big\langle G'(t)(u(t) - u^\varepsilon(t)), v(t)\big\rangle + \big\langle G'(t)u^\varepsilon(t), v(t) - v^\varepsilon(t)\big\rangle\big]\psi(t)\big|\,dt \\ &\leqslant \|G'\|_{L^\infty(I;\mathcal{L}(X,Y^*))}\|\psi\|_{L^\infty(I)}\Big[\|u - u^\varepsilon\|_{L^2(I;X)}\|v\|_{L^2(I;Y)} \\ &\qquad\qquad + \|v - v^\varepsilon\|_{L^2(I;Y)}\|u^\varepsilon\|_{L^2(I;X)}\Big].\end{aligned}$$

The right-hand side converges to zero when $\varepsilon \to 0$. The other expressions can be handled in the same way, which proves the assertion. □

We refer to the last lemma as a *product rule*, because when choosing v as constant in time it states that $(\mathcal{G}u)' = \mathcal{G}'u + \mathcal{G}u'$, where $\mathcal{G}'$ denotes the realization of G'.

Before continuing with the discussion under which circumstances (2.2) is solvable, we give a short example of how to restate actual partial differential equations as evolution equations. Although we will see far more comprehensive examples for this in Chapters 5 and 6, we consider it crucial for understanding the consequences of the assumptions on the operators and the data. Only then are we able to judge which assumptions are acceptable and which are too excessive. To keep the example brief we use the wave equation with a single time- and space-dependent coefficient for this demonstration.

Example 2.4. Consider the initial boundary value problem

$$u''(t,x) - \operatorname{div}(a(t,x)\nabla u(t,x)) = f(t,x) \quad t \in (0,T),\ x \in \Omega$$
$$u = 0 \text{ on } (0,T) \times \partial\Omega,\ u(0,\cdot) = u_0,\ u'(0,\cdot) = u_1$$

on some domain $\Omega \subset \mathbb{R}^d$ (with $d \in \mathbb{N}$). A weak formulation of this wave equation is easily obtained and reads

$$\int_\Omega u''(t,x)\varphi(x)\,dx + a(t,x)\nabla u(t,x) \cdot \nabla\varphi(x)\,dx = \int_\Omega f(t,x)\varphi(x)\,dx,$$

and should be fulfilled for all $\varphi \in H_0^1$ and almost all $t \in I$. Typically, one seeks a solution $u(t) \in H_0^1(\Omega)$ that possesses a time derivative $u'(t) \in L^2(\Omega)$, thus $V = H_0^1(\Omega)$ and $H = L^2(\Omega)$. In effect, $V^* = H^{-1}(\Omega)$. According to Lemma 2.2, this weak formulation is equivalent to $(\mathcal{C}u')' + \mathcal{A}u = f$ in $L^2(I; H^{-1}(\Omega))$, where $C(t) := \mathrm{Id}_{L^2(\Omega)}$ and

$$\langle A(t)\varphi, \psi\rangle := \int_\Omega a(t,x)\ \nabla\varphi(x) \cdot \nabla\psi(x)\,dx, \quad \varphi, \psi \in H_0^1(\Omega)$$

for almost all $t \in I$. We see that smoothness assumptions on A directly translate to the coefficient a. In particular, we have $A \in L^\infty(I; \mathcal{L}(V, V^*))$ if and only if $a \in L^\infty(I; L^\infty(\Omega))$.

2.2 EXISTENCE OF SOLUTIONS

We use the Galerkin method in a variational context to show well-posedness of the solution $u \in L^2(I;V) \cap H^1(I;H)$ of (2.2) with the initial conditions that we discussed in the previous section, i.e. the problem

$$(\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{A} + \mathcal{Q})u = f \quad \text{in } L^2(I;V^*), \tag{2.6a}$$
$$u(0) = u_0 \quad \text{and} \quad (\mathcal{C}u')(0) = u_1. \tag{2.6b}$$

Although this should be fairly standard, the literature lacks a satisfying presentation of the needed results. For instance, the popular works by Lions and Magenes [LM72] and Zeidler [Zei90b] only consider the case of time-independent A, $B = Q = 0$ and $C = \mathrm{Id}$, as does Evans [Eva10]. The extension to time-dependent $A = A(t)$ can be found in [Wlo87], and the special case that A represents the operator $-\Delta + c(t)$ has been analyzed by us in [GL17]. Lions [Lio61] analyzes our equation by extending the equation from $(0,T)$ to $(0,\infty)$ and then applying a generalized Lax-Milgram-Lemma to the problem's space-time formulation. We will encounter this technique in Section 3.1, but it has the drawback of requiring homogeneous initial values. The results of Lions and Dautray [DL00] are close to what we need, but their technique (parabolic regularization) does not yield continuous dependence of u on the operators A, B, C and Q. However, this knowledge is crucial in the analysis of the inverse problems that will appear in subsequent chapters.

Another point that will turn out to be important to us, but is missing in these results, is the admissibility of right-hand sides $f(t) \in V^*$. Consequently, we will generalize the known results to (2.6) ourselves. Our approach is similar to the one of [LM72].

We start by showing existence of a solution u using the Galerkin method. This means that we approximate the evolution equation (2.6) and the initial conditions by restricting the test- and trial functions to finite-dimensional subspaces V_m of V and then let the dimension of these subspaces tend to infinity.

We assumed V to be separable, so it possesses a Schauder basis on which we can apply the Gram-Schmidt orthonormalization algorithm. Since V is dense in H, this procedure yields a complete orthonormal system $(\varphi_j)_{j\in\mathbb{N}} \subset V$ of H. The subspaces of V in which we will seek approximative solutions u_m are $V_m := \operatorname{lin}\{\, \varphi_j \mid j = 1, \dots, m \,\}$. Showing existence of these u_m is relatively easy, because due to the finite dimension of V_m the resulting problem can be reduced to a linear system of ordinary differential equations. To make it solvable, we have to ensure that it can be converted to an explicit system.

Theorem 2.5. *Let $m \in \mathbb{N}$ and $A \in L^\infty(I; \mathcal{L}(V, V^*))$, $B \in L^\infty(I; \mathcal{L}(H))$, $C \in W^{1,\infty}(I; \mathcal{L}(H))$ and $Q \in L^\infty(I; \mathcal{L}(V, H))$. Further, let $u_0, u_1 \in H$, $f \in L^2(I; V^*)$ and let C be uniformly pointwise H-coercive, i.e.*

$$(C(t)v, v) \geqslant c_0 \|v\|_H^2 \quad \textit{for all } v \in H$$

holds for almost all $t \in I$ for some $c_0 > 0$ (that does not depend on t). Then there exists a unique solution $u_m \in H^2(I; V_m)$ of the equation

$$(\mathcal{C}u_m')' + \mathcal{B}u_m' + (\mathcal{A} + \mathcal{Q})u_m = f \quad \textit{in } L^2(I; V_m^*) \tag{2.7a}$$

that also satisfies the initial conditions

$$u_m(0) = \sum_{j=1}^{m} (u_0, \varphi_j)\varphi_j \quad \textit{and} \quad u_m'(0) = \sum_{j=1}^{m} \Big(C(0)^{-1}u_1, \varphi_j\Big)\varphi_j. \tag{2.7b}$$

Proof. First, we note that C is continuous in time (due to Theorem A.10). As a consequence, $C(t)$ is coercive (and therefore invertible) for *all* $t \in I$. In particular, the second initial condition in (2.7b) is well-defined.

The searched for $u_m \in H^2(I; V_m)$ has the shape $u_m(t) = \sum_{j=1}^m \alpha_j(t)\varphi_j$ for almost all $t \in I$, with $\alpha_j \in H^2(I)$ for $j = 1, \dots, m$. Instead of u_m, we can equivalently search for these coefficient functions. Equation (2.7a) holds if and only if

$$\sum_{j=1}^{m} \bigg(\frac{d}{dt}\big(\alpha_j'(t)\big(C(t)\varphi_j, \varphi_i\big)\big) + \alpha_j'(t)\big(B(t)\varphi_j, \varphi_i\big) \\ + \alpha_j(t)\big\langle (A(t) + Q(t))\varphi_j, \varphi_i \big\rangle \bigg) = \langle f(t), \varphi_i \rangle$$

is satisfied for almost all $t \in I$ and all $i = 1, \dots, m$. We note that $\alpha_j' \in H^1(I)$ and $\big(C(\cdot)\varphi_j, \varphi_i\big) \in W^{1,\infty}(I)$ (special case of Lemma 2.3). Hence, we can apply the product rule for weak derivatives to obtain the equivalent system

$$M_2(t) \cdot \alpha''(t) + M_1(t) \cdot \alpha'(t) + M_0(t) \cdot \alpha(t) = F(t) \tag{2.8}$$

of explicit second-order ODEs for $\alpha = (\alpha_1, \dots, \alpha_m)$, with $m \times m$ matrices

$$\begin{aligned} M_2(t) &:= \Big(\big(C(t)\varphi_j, \varphi_i\big)\Big)_{i,j=1,\dots,m} \\ M_1(t) &:= \Big(\big(C'(t)\varphi_j, \varphi_i\big) + \big(B(t)\varphi_j, \varphi_i\big)\Big)_{i,j=1,\dots,m} \\ M_0(t) &:= \Big(\big\langle A(t)\varphi_j, \varphi_i\big\rangle + \big(Q(t)\varphi_j, \varphi_i\big)\Big)_{i,j=1,\dots,m} \end{aligned}$$

and the right-hand side vector $F(t) := (\langle f(t), \varphi_i\rangle)_{i=1,\dots,m}$. Due to Lax-Milgram we know that $M_2(t)$ is invertible with $\left\|M_2(t)^{-1}\right\| \leqslant c_0^{-1}$. We conclude that

$$M_2(\cdot)^{-1} M_1(\cdot), \quad M_2(\cdot)^{-1} M_0(\cdot) \in L^\infty(I; \mathbb{R}^{m \times m})$$

and $M_2(\cdot)^{-1}F(\cdot) \in L^2(I; \mathbb{R}^m)$. The initial values that u_m is required to possess attain the shape

$$\begin{aligned} \alpha(0) &= ((u_0, \varphi_i))_{i=1,\dots,m}, \\ \alpha'(0) &= \left(\left\langle C(0)^{-1} u_1, \varphi_i \right\rangle\right)_{i=1,\dots,m}. \end{aligned} \tag{2.9}$$

According to (weak) ODE theory (e.g. [CL55; ORe97]), the initial value problem (2.8) and (2.9) admits a unique solution $\alpha \in H^2(I; \mathbb{R}^m)$. □

We would like to highlight the fact that the existence of u_m does not depend on coercivity of A. This is a side effect of V_m being finite-dimensional, because this makes the H- and V-norm on V_m equivalent. Hence, we did not need A to ensure that the solution u_m belongs not only in H, but also to V. This is going to change now, because we require estimates for u_m that are independent of m in order to obtain a limit when we let m tend to infinity.

To prove these estimates we will employ the following version of the *Grönwall*-Lemma, taken from [DL00].

Lemma 2.6 (Grönwall). *Let* $\varphi \in L^\infty(I), \mu \in L^1(I)$ *satisfy* $\varphi(t) \geqslant 0$, $\mu(t) \geqslant 0$ *and the integral inequality*

$$\varphi(t) \leqslant \int_0^t \mu(s)\varphi(s)\, ds + C \tag{2.10}$$

for almost all $t \in I$ *with* $C > 0$. *Then*

$$\varphi(t) \leqslant C \exp\left(\int_0^t \mu(s)\, ds\right) \tag{2.11}$$

also holds for almost all $t \in I$.

The important thing to note here is that in the integral inequality (2.10) φ appears on both sides, whereas the right-hand side of (2.11) only depends on μ. The proof is even shorter than the assertion, therefore we include it.

Proof. Define $F(t) := \int_0^t \mu(s)\varphi(s)\,ds + C$, then $F'(t) = \mu(t)\varphi(t)$ and (2.10) states $\frac{F'(t)}{F(t)} \leqslant \mu(t)$. By integrating this we obtain

$$\log\left(\frac{F(t)}{C}\right) \leqslant \int_0^t \mu(s)\,ds,$$

which means $\varphi(t) \leqslant F(t) \leqslant C\exp\left(\int_0^t \mu(s)\,ds\right)$. □

For the estimates we will not only need coercivity of $A(t)$, but also symmetry of $A(t)$ and $C(t)$, i.e. $\langle A(t)\varphi, \psi\rangle = \langle A(t)\psi, \varphi\rangle$ and $(C(t)u, v) = (C(t)v, u)$ for all $\varphi, \psi \in V$ and $u, v \in H$. This means that $C(t)$ should be self-adjoint. The adjoint of $A(t) \in \mathcal{L}(V, V^*)$ is an element of $\mathcal{L}(V^{**}, V^*)$, but since V is reflexive we can identify V^{**} with V and therefore regard $A^*(t)$ as an operator $\mathcal{L}(V, V^*)$ as well, which is then determined by the familiar relation $\langle A^*(t)\varphi, \psi\rangle = \langle A(t)\psi, \varphi\rangle$. To ease notation we make the following definitions.

Definition 2.7. Let Z be a reflexive Banach space and $\alpha > 0$. We introduce the set

$$\mathcal{L}^{sa}(Z, Z^*) := \left\{ G \in \mathcal{L}(Z, Z^*) \;\middle|\; G^* = G \right\}.$$

of self-adjoint operators mapping from Z to Z^*. Its subset of α-coercive self-adjoint operators we denote by

$$\mathcal{L}^{sa}_\alpha(Z, Z^*) := \left\{ G \in \mathcal{L}^{sa}(Z, Z^*) \;\middle|\; \langle Gz, z\rangle \geqslant \alpha\|z\|_Z^2 \text{ for all } z \in Z \right\}.$$

If $Z = Z^*$, then we abbreviate $\mathcal{L}^{sa}(Z, Z^*)$ and $\mathcal{L}^{sa}_\alpha(Z, Z^*)$ by $\mathcal{L}^{sa}(Z)$ and $\mathcal{L}^{sa}_\alpha(Z)$, respectively.

These sets will also ease the analysis of the inverse problems in subsequent chapters because they can be used to construct the domain of definition of the map $(A, C) \mapsto u$. Note that $\mathcal{L}^{sa}(Z, Z^*)$ is a closed subspace of $\mathcal{L}(Z, Z^*)$, i.e. $\mathcal{L}^{sa}(Z, Z^*)$ is a Banach space when it is equipped with the operator norm. Thus we can also use the corresponding Bochner spaces $W^{k,p}(I; \mathcal{L}^{sa}(Z, Z^*))$. Since $\mathcal{L}^{sa}_\alpha(Z, Z^*)$ is not even a linear space, we must however refrain from writing $W^{k,p}(I; \mathcal{L}^{sa}_\alpha(Z, Z^*))$.

Theorem 2.8. *Let $a_0, c_0 > 0$ and suppose that $A \in W^{1,\infty}(I; \mathcal{L}^{sa}(V, V^*))$, $B \in L^\infty(I; \mathcal{L}(H))$, $Q \in L^\infty(I; \mathcal{L}(V, H))$ and $C \in W^{1,\infty}(I; \mathcal{L}^{sa}(H))$ with $A(t) \in \mathcal{L}^{sa}_{a_0}(V, V^*)$ and $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$ for almost all $t \in I$. Further, let $u_0 \in V$, $u_1 \in H$ and $f \in L^2(I; H) \cup H^1(I; V^*)$.*
For every $m \in \mathbb{N}$ we have the estimate

$$\operatorname*{ess\,sup}_{t\in I} \|u_m(t)\|_V^2 + \|u_m'(t)\|_H^2 \leqslant \Lambda\left(\|u_0\|_V^2 + \|u_1\|_H^2 + \|f\|^2\right)$$

for the solution u_m *of problem* (2.7). *With* $\|f\|$ *we denote the appropriate norm of* f*, i.e. either its* $L^2(I;H)$*- or* $H^1(I;V^*)$*-norm. The constant* $\Lambda > 0$ *continuously depends on* a_0^{-1}, c_0^{-1} *and the norms of the operators* A, B, C *and* Q.

Proof. We can do here what we are not allowed to do in the original equation, namely testing the equation (2.7a) with $u_m'(t) \in V$. This results in

$$\begin{aligned}\langle f(t), u_m'(t)\rangle = \Big\langle (\mathcal{C}u_m')'(t), u_m'(t)\Big\rangle + \big(B(t)u_m'(t), u_m'(t)\big) \\ + \big\langle A(t)u_m(t), u_m'(t)\big\rangle + \big(Q(t)u_m(t), u_m'(t)\big).\end{aligned} \tag{2.12}$$

Our goal is to reshape this until we obtain an inequality like (2.10) that is needed for Grönwall's Lemma with $\varphi(t) = \|u_m(t)\|_V^2 + \|u_m'(t)\|_H^2$. In order to be able to benefit from the coercivity of $A(t)$ and $C(t)$, we apply the product rule of Lemma 2.3 to the corresponding terms to obtain

$$2\big\langle A(t)u_m(t), u_m'(t)\big\rangle = \frac{d}{dt}\langle A(t)u_m(t), u_m(t)\rangle - \big\langle A'(t)u_m(t), u_m(t)\big\rangle$$

for A and analogously

$$\begin{aligned}2\Big(\big(\mathcal{C}u_m'\big)'(t), u_m'(t)\Big) &= 2\big(C(t)u_m''(t), u_m'(t)\big) + 2\big(C'(t)u_m'(t), u_m'(t)\big) \\ &= \frac{d}{dt}\big(C(t)u_m'(t), u_m'(t)\big) + \big(C'(t)u_m'(t), u_m'(t)\big)\end{aligned}$$

for the expression that involves C. By plugging these equalities into (2.12) and integrating over $(0, t)$ we reach

$$\begin{aligned}&\big(C(t)u_m'(t), u_m'(t)\big) + \langle A(t)u_m(t), u_m(t)\rangle \\ &\quad= \big(C(0)u_m'(0), u_m'(0)\big) + \langle A(0)u_m(0), u_m(0)\rangle \\ &\qquad- \int_0^t 2\big(B(s)u_m'(s), u_m'(s)\big) + \big(C'(s)u_m'(s), u_m'(s)\big)\,ds \\ &\qquad+ \int_0^t \big\langle A'(s)u_m(s), u_m(s)\big\rangle - 2\big(Q(s)u_m(s), u_m'(s)\big)\,ds \\ &\qquad+ \int_0^t 2\big\langle f(s), u_m'(s)\big\rangle\,ds.\end{aligned} \tag{2.13}$$

Here we used the fundamental theorem of calculus for $H^1(I)$ and the fact that A and C are continuous in time. These are consequences of Theorem A.10.

Equation (2.13) can be viewed as a *energy equality* for u_m: The energy at time t is determined by the initial energy and the effects of the right-hand side, the operators Q and B and changes in A and C since then. The term *energy* stems from the fact that e.g. for the classical wave equation $u'' - \Delta u = f$ the left-hand side would read $\|u_m'(t)\|_H^2 + \|\nabla u_m(t)\|_H^2$, which is the sum of kinetic and potential energy of $u(t)$.

The left-hand side of (2.13) can be estimated using the coercivity of A and C, which yields

$$\big(C(t)u_m'(t), u_m'(t)\big) + \langle A(t)u_m(t), u_m(t)\rangle \geqslant c_0\big\|u_m'(t)\big\|_H^2 + a_0\big\|u_m(t)\big\|_V^2.$$

The analogous terms for the initial values on the right-hand side are bounded from above by

$$\begin{aligned}&\big(C(0)u_m'(0), u_m'(0)\big) + \langle A(0)u_m(0), u_m(0)\rangle \\ &\quad\leqslant \|C(0)\|_{\mathcal{L}(H)}\|C(0)^{-1}u_1\|_H^2 + \|A(0)\|_{\mathcal{L}(V,V^*)}\|u_0\|_V^2 \\ &\quad\leqslant c_0^{-2}\|C\|_{L^\infty(I;\mathcal{L}(H))}\|u_1\|_H^2 + \|A\|_{L^\infty(I;\mathcal{L}(V,V^*))}\|u_0\|_V^2.\end{aligned}$$

Terms in the integral of the right-hand side of (2.13) that involve A, B or C are easily estimated using their operator norms. For Q we note that

$$\begin{aligned}2\big(Q(s)u_m(s), u_m'(s)\big) &\leqslant 2\|Q\|_{L^\infty(I;\mathcal{L}(V,H))}\|u_m(s)\|_V\|u_m'(s)\|_H \\ &\leqslant \|Q\|_{L^\infty(I;\mathcal{L}(V,H))}\Big(\|u_m(s)\|_V^2 + \|u_m'(s)\|_H^2\Big).\end{aligned}$$

The treatment of $f \in L^2(I;H)$ is easily done because $2\langle f(s), u_m'(s)\rangle \leqslant \|f(s)\|_H^2 + \|u_m'(s)\|_H^2$. Otherwise, i.e. $f \in H^1(I;V^*)$, we have for all $\varepsilon > 0$ the estimate

$$\begin{aligned}2\int_0^t \big\langle f(s), u_m'(s)\big\rangle\, ds &= 2\langle f(t), u_m(t)\rangle - 2\langle f(0), u_m(0)\rangle \\ &\qquad - 2\int_0^t \big\langle f'(s), u_m(s)\big\rangle\, ds \\ &\leqslant \varepsilon\|u_m(t)\|_V^2 + \tfrac{1+\varepsilon}{\varepsilon}\|f\|_{L^\infty(I;V^*)}^2 + \|u_0\|_V^2 \\ &\qquad + \int_0^t \big\|f'(s)\big\|_{V^*}^2 + \|u_m(s)\|_V^2\, ds.\end{aligned}$$

The norm $\|f\|_{L^\infty(I;V^*)}$ can be estimated by $C_{H^1\hookrightarrow L^\infty}\|f\|_{H^1(I;V^*)}$, where $C_{H^1\hookrightarrow L^\infty} > 0$ denotes the norm of the embedding $H^1(I;V^*) \hookrightarrow L^\infty(I;V^*)$. We can combine both cases because

$$\begin{aligned}2\int_0^t \big\langle f(s), u_m'(s)\big\rangle\, ds &\leqslant \varepsilon\|u_m(t)\|_V^2 + \big(\varepsilon^{-1} + C_{H^1\hookrightarrow L^\infty}^2 + 1\big)\|f\|^2 + \|u_0\|_V^2 \\ &\qquad + \int_0^t \big\|u_m(s)\big\|_V^2 + \big\|u_m'(s)\big\|_H^2\, ds\end{aligned}$$

holds for either form of f as long as $\|f\|$ denotes the norm of f in the appropriate space.

By inserting all of the above considerations into (2.13) we obtain

$$\begin{aligned}&c_0\big\|u_m'(t)\big\|_H^2 + (a_0 - \varepsilon)\|u_m(t)\|_V^2 \\ &\quad\leqslant \|C\|c_0^{-2}\|u_1\|_H^2 + (1 + \|A\|)\|u_0\|_V^2 + \big(\varepsilon^{-1} + C_{H^1\hookrightarrow L^\infty} + 1\big)\|f\|^2 \\ &\qquad + \int_0^t (1 + 2\|B\| + \big\|C'\big\| + \|Q\|)\big\|u_m'(s)\big\|_H^2\, ds \\ &\qquad + \int_0^t (1 + \big\|A'\big\| + \|Q\|)\|u_m(s)\|_V^2\, ds,\end{aligned}$$

where we abbreviated the appropriate L^∞-norms of the operators for typographic reasons. Here it becomes clear that to allow $f(t) \in V^*$ we have to sacrifice some coercivity on the left-hand side, therefore possibly worsening the estimate. Fortunately, we are not interested in a sharp upper bound for the norm of u_m; we are satisfied as long as this bound continuously depends on the data u_0, u_1, f and the operators A, B, C, Q. Hence we can just choose e.g. $\varepsilon := a_0/2$ and note that there exists a positive constant λ such that

$$\begin{aligned} &\|u_m'(t)\|_H^2 + \|u_m(t)\|_V^2 \\ &\quad \leqslant \lambda\Big(\|u_0\|_V^2 + \|u_1\|_H^2 + \|f\|^2\Big) + \lambda \int_0^t \|u_m'(s)\|_H^2 + \|u_m(s)\|_V^2 \,ds \end{aligned} \tag{2.14}$$

holds for all $t \in I$. Further, λ continuously depends on the operators (measured in the spaces given in the assertion) as well as on c_0^{-1} and a_0^{-1}. In particular, λ stays bounded as long as these values stay bounded. We are finally in a position to apply Grönwall's Lemma, which yields

$$\|u_m'(t)\|_H^2 + \|u_m(t)\|_V^2 \leqslant \lambda \exp(\lambda T)\Big(\|u_0\|_V^2 + \|u_1\|_H^2 + \|f\|^2\Big).$$

Forming the supremum over I of both sides of this equation and setting $\Lambda := \lambda \exp(\lambda T)$ concludes the proof. □

We would like to point out that it would suffice if the operator A was only coercive up to a perturbation in H, i.e. satisfied $\langle A(t)\varphi, \varphi\rangle \geqslant a_0\|\varphi\|_V^2 - \lambda_0\|\varphi\|_H^2$ with $\lambda_0 \in \mathbb{R}$, often labeled as *Gårding's inequality*. Because we included the operator Q, we could just replace such an A with $A + \lambda_0 I$ and Q with $Q - \lambda_0 I$. This does not change u_m, but causes $A(t)$ to be coercive.

Next, we perform the limit $m \to \infty$.

Theorem 2.9. *Let $a_0, c_0 > 0$ and suppose that $A \in W^{1,\infty}(I;\mathcal{L}^{sa}(V,V^*))$, $B \in L^\infty(I;\mathcal{L}(H))$, $Q \in L^\infty(I;\mathcal{L}(V,H))$ and $C \in W^{1,\infty}(I;\mathcal{L}^{sa}(H))$ with $A(t) \in \mathcal{L}^{sa}_{a_0}(V,V^*)$ and $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$ for almost all $t \in I$. Further, let $u_0 \in V$, $u_1 \in H$ and $f \in L^2(I;H) \cup H^1(I;V^*)$.*
There exists a solution $u \in L^2(I;V) \cap H^1(I;H)$ to problem (2.6) that also satisfies the energy estimate

$$\operatorname*{ess\,sup}_{t\in I} \|u(t)\|_V^2 + \|u'(t)\|_H^2 \leqslant \Lambda\left(\|u_0\|_V^2 + \|u_1\|_H^2 + \|f\|^2\right).$$

With $\|f\|$ we again denote the appropriate norm of f. The constant $\Lambda > 0$ continuously depends on a_0^{-1}, c_0^{-1} and the operators A, B, C and Q.

Proof. From the energy estimates for u_m we conclude that the sequence $(u_m)_{m\in\mathbb{N}}$ is bounded in the Hilbert space $L^2(I;V)$ and $(u_m')_{m\in\mathbb{N}}$ is bounded in $L^2(I;H)$. Therefore they possess weakly convergent subsequences, i.e. there exist $u \in L^2(I;V)$, $v \in L^2(I;H)$ with

$$\begin{aligned} u_{m_j} &\rightharpoonup u \quad \text{in } L^2(I;V), \\ u_{m_j}' &\rightharpoonup v \quad \text{in } L^2(I;H). \end{aligned}$$

as $j \to \infty$. Without loss of generality we assume that the indices of both sequences are identical. The fact that $v = u'$ can for example be seen by noting that $(u_m)_{m\in\mathbb{N}}$ also has a subsequence that converges weakly to $w \in H^1(I;H)$. Due to uniqueness of weak limits we have $w = u$ and $w' = v$, both in the sense of $L^2(I;H)$.

By means of $\|u\| \leqslant \liminf_{j\to\infty} \|u_{m_j}\|$ we obtain the L^2-energy estimates

$$\|u\|^2_{L^2(I;V)} + \|u'\|^2_{L^2(I;H)} \leqslant \Lambda\Big(\|u_0\|^2_V + \|u_1\|^2_H + \|f\|^2\Big)$$

for u. We are also able to achieve L^∞-estimates in time, because $(u_m)_{m\in\mathbb{N}}$ is bounded in $L^\infty(I;V) \cong L^1(I;V^*)^*$, while the sequence $(u'_m)_{m\in\mathbb{N}}$ is bounded in $L^\infty(I;H) \cong L^1(I;H)^*$. In this case the Banach–Alaoglu theorem states that there are a subsequences

$$\begin{aligned} u_{m_j} &\overset{*}{\rightharpoonup} u \quad \text{in } L^1(I;V^*),\\ u'_{m_j} &\overset{*}{\rightharpoonup} u' \quad \text{in } L^1(I;H) \end{aligned}$$

when $j \to \infty$. The limits are the same as before: For example, weak-$*$ convergence of $(u_{m_j})_{j\in\mathbb{N}}$ in $L^1(I;V^*)$ is stronger than weak convergence in $L^2(I;V)$. The former is the case if

$$\int_0^T \big\langle v(t), u_{m_j}(t)\big\rangle\,dt \to \int_0^T \langle v(t), u(t)\rangle\,dt$$

holds for all $v \in L^1(I;V^*)$, while the latter only requires this to hold for $v \in L^2(I;V^*)$. Hence, u satisfies the same type of energy estimates as u_m.

We still need to show that u indeed solves (2.6). To this end, let $\psi \in C_c^\infty(I)$ and $\varphi \in \operatorname{lin}\{\,\varphi_i \mid i \in \mathbb{N}\,\}$. Thus, there exists $N \in \mathbb{N}$ such that

$$\begin{aligned}\int_0^T \langle f(t),\varphi\rangle\psi(t)\,dt = \int_0^T &\Big\langle B(t)u'_{m_j}(t) + (A(t)+Q(t))u_{m_j}(t), \varphi\Big\rangle\psi(t)\\ &- \Big(C(t)u'_{m_j}(t),\varphi\Big)\psi'(t)\,dt\end{aligned}$$

holds for all $j \geqslant N$. The right-hand side is a continuous linear form in $u_{m_j} \in L^2(I;V)\cap H^1(I;H)$. Due to the weak convergence of $u_{m_j} \to u$ in this space we conclude that the formulation also holds for u, that is

$$\begin{aligned}\int_0^T \langle f(t),\varphi\rangle\psi(t)\,dt = \int_0^T &\langle B(t)u'(t) + (A(t)+Q(t))u(t), \varphi\rangle\psi(t)\\ &- (C(t)u'(t),\varphi)\psi'(t)\,dt\end{aligned}$$

Similarly, both sides of this equation are linear in φ and continuous with respect to its norm in V. Since $\operatorname{lin}\{\,\varphi_i \mid i \in \mathbb{N}\,\}$ is dense in V the equation thus holds for all $\varphi \in V$. According to Lemma 2.2, this shows that u solves evolution equation (2.6a).

Now we turn to the initial values (2.6b). Let $N \in \mathbb{N}$ and $\beta \in C^2(I;V)$ with $\beta(t) \in \operatorname{lin}\{\,\varphi_i \mid i = 1,\dots,N\,\}$ for all $t \in I$ and $\beta(T) = \beta'(T) = 0$. For $m_j \geqslant N$, we may test the weak formulation for u_{m_j} with $\beta(t)$ and integrate over the time interval. By doing so we obtain

$$\begin{aligned}&\int_0^T \Big\langle (\mathcal{C}u'_{m_j})'(t), \beta(t)\Big\rangle\,dt\\ &\quad= \int_0^T \Big\langle f(t) - B(t)u'_{m_j}(t) - (A(t)+Q(t))u_{m_j}(t), \beta(t)\Big\rangle\,dt,\end{aligned} \tag{2.15}$$

and we can derive an analogous equation for u. The right-hand side of equation (2.15) converges to the respective term for u for $j \to \infty$, therefore the left-hand sides have to converge as well, i.e.

$$\int_0^T \Big\langle (\mathcal{C}u'_{m_j})'(t), \beta(t) \Big\rangle \, dt \to \int_0^T \Big\langle (\mathcal{C}u')'(t), \beta(t) \Big\rangle \, dt. \tag{2.16}$$

We apply the integration by parts formula in the evolution triple $H \subset V^* \subset H^*$ (see Theorem A.11), starting with the right-hand side of (2.16). By doing so twice and considering the boundary values of β at $t = T$ and the fact that $C(t)$ is self-adjoint, we see

$$\begin{aligned}\int_0^T \Big\langle (\mathcal{C}u')'(t), \beta(t) \Big\rangle \, dt &= -\int_0^T \big(C(t)u'(t), \beta'(t)\big) \, dt - \big\langle (\mathcal{C}u')(0), \beta(0) \big\rangle \\ &= \int_0^T \Big(u(t), (\mathcal{C}\beta')'(t)\Big) \, dt - \big\langle (\mathcal{C}u')(0), \beta(0) \big\rangle + \big(u(0), (\mathcal{C}\beta')(0)\big).\end{aligned} \tag{2.17}$$

The same argument can be made for the left-hand side of (2.16), which means

$$\begin{aligned}&\int_0^T \Big\langle (\mathcal{C}u'_{m_j})'(t), \beta(t) \Big\rangle \, dt \\ &\quad = \int_0^T \Big(u_{m_j}(t), (\mathcal{C}\beta')'(t)\Big) \, dt - \Big((\mathcal{C}u'_{m_j})(0), \beta(0)\Big) + \big(u_{m_j}(0), (\mathcal{C}\beta')(0)\big).\end{aligned}$$

Now we insert the initial values (2.7b) of u_{m_j} and thereby obtain

$$\big(u_{m_j}(0), (\mathcal{C}\beta')(0)\big) = \left(\sum_{i=1}^{m_j} (u_0, \varphi_i)\varphi_i, (\mathcal{C}\beta')(0)\right),$$

which clearly converges to $(u_0, (\mathcal{C}\beta')(0))$ when $j \to \infty$. Moreover, from $u_{m_j} \in H^2(I; V)$ follows that u'_{m_j} is continuous (with values in V), hence

$$\Big((\mathcal{C}u'_{m_j})(0), \beta(0)\Big) = \Big(C(0)u'_{m_j}(0), \beta(0)\Big) = \Big(u'_{m_j}(0), C(0)\beta(0)\Big).$$

Due to the design of the second initial condition for the Galerkin solutions this tends to $\Big(C(0)^{-1}u_1, C(0)\beta(0)\Big)$, which is equal to $(u_1, \beta(0))$. Since $\int_0^T \Big\langle (\mathcal{C}u'_{m_j})'(t), \beta(t) \Big\rangle$ also converges to the right-hand side of (2.17), we are able to conclude

$$\big(u(0), (\mathcal{C}\beta')(0)\big) - \big\langle (\mathcal{C}u')(0), \beta(0) \big\rangle = \big(u_0, (\mathcal{C}\beta')(0)\big) - (u_1, \beta(0)).$$

By an approximation argument this holds for all $\beta \in C^2(I; V)$ with $\beta(T) = \beta'(T) = 0$. Since $C(0)$ is invertible, the initial values $\beta(0)$ and $(\mathcal{C}\beta')(0)$ can be arbitrarily chosen in V and H, respectively. Thus, $u(0) = u_0$ in H and $(\mathcal{C}u')(0) = u_1$ in V^*. This concludes the proof. □

2.3 UNIQUENESS OF SOLUTIONS

We also need u to be unique. According to [DL00] it is not known whether the assumptions we made to get existence of a solution also ensure its uniqueness. The standard proof, which is widely used in literature (e.g. in [DL00; LM72; Wlo87]) and which we adapted to (2.6), additionally requires differentiability of B in time. In the case that Q indeed represents a first-order differential operator in space, i.e. $Q \in L^\infty(I;\mathcal{L}(V,H)) \setminus L^\infty(I;\mathcal{L}(H))$, we also need it to be differentiable. For convenience we combine the uniqueness with the already proven existence result.

Theorem 2.10. *Let $a_0, c_0 > 0$ and suppose that $A \in W^{1,\infty}(I;\mathcal{L}^{sa}(V,V^*))$, $B \in W^{1,\infty}(I;\mathcal{L}(H))$ and $C \in W^{1,\infty}(I;\mathcal{L}^{sa}(H))$ with $A(t) \in \mathcal{L}^{sa}_{a_0}(V,V^*)$ and $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$ for almost all $t \in I$. Further, let $Q \in W^{1,\infty}(I;\mathcal{L}(V,H))$ or $Q \in L^\infty(I;\mathcal{L}(H))$, $u_0 \in V$, $u_1 \in H$ and $f \in L^2(I;H) \cup H^1(I;V^*)$.*
Then there exists a unique solution $u \in L^2(I;V) \cap H^1(I;H)$ to problem (2.6). Furthermore, this solution satisfies the energy estimate

$$\operatorname*{ess\,sup}_{t\in I} \|u(t)\|_V^2 + \|u'(t)\|_H^2 \leqslant \Lambda\left(\|u_0\|_V^2 + \|u_1\|_H^2 + \|f\|^2\right).$$

With $\|f\|$ we denote either the norm of f in $L^2(I;H)$ or in $H^1(I;V^*)$, whichever is finite. The constant $\Lambda > 0$ continuously depends on a_0^{-1}, c_0^{-1} and the operators A, B, C and Q, measured in $W^{1,\infty}(I;\mathcal{L}(V,V^*))$, $L^\infty(I;\mathcal{L}(H))$, $W^{1,\infty}(I;\mathcal{L}(H))$ and $L^\infty(I;\mathcal{L}(V,H))$, respectively.

Proof. Existence of a solution u that satisfies the energy estimates is known from Theorem 2.9. Due to linearity of u in the tuple $(f, u_0, u_1) \in L^2(I;V^*) \times V \times H$ it is sufficient to show that $u = 0$ is the only solution to the problem with homogeneous data $f = u_0 = u_1 = 0$. We are going to derive estimates for u that show that is has to vanish at all $t \in I$. For the Galerkin solutions u_m we were able to test the evolution equation with u_m' to infer the energy estimates, but this is not permitted for u itself.[3] Hence, we go "one level higher" and test the equation with an anti-derivative of u. This will result in an integral inequality on which we can apply Grönwall's Lemma once again, to obtain an estimate for u and $\int_0^t u(s)\,ds$ (instead of u' and u). Precisely, for $s \in I$ we define the function $\psi_s \in H^1(I;V)$ via

$$\psi_s(t) = \begin{cases} -\int_t^s u(\sigma)\,d\sigma & \text{if } t < s, \\ 0 & \text{otherwise.} \end{cases}$$

After testing the equation with $\psi_s(t)$ and integrating over $t \in I$, we may apply the integration by parts formula on the expression that involves C. Keeping in mind that $u_1 = 0$ and $\psi_s(t) = 0$ for $t \geqslant s$, we arrive at

$$0 = \int_0^s \big\langle B(t)u'(t) + [A(t) + Q(t)]u(t), \psi_s(t)\big\rangle - \big(C(t)u'(t), \psi_s'(t)\big)\,dt. \tag{2.18}$$

3 This is a major difference compared to the parabolic case: There it is enough to test the evolution equation with $u(t)$. Thus, the energy estimates hold for *all* solutions, and uniqueness follows immediately.

Invoking the product rule of Lemma 2.3 yields

$$2\big(C(t)u'(t), \psi_s'(t)\big) = 2\big(C(t)u'(t), u(t)\big) = \frac{d}{dt}(C(t)u(t), u(t)) - \big(C'(t)u(t), u(t)\big)$$

for almost all $t \in I$, as well as

$$2\langle A(t)u(t), \psi_s(t)\rangle = 2\big\langle A(t)\psi_s'(t), \psi_s(t)\big\rangle = \frac{d}{dt}\langle A(t)\psi_s(t), \psi_s(t)\rangle - \big\langle A'(t)\psi_s(t), \psi_s(t)\big\rangle$$

because both operators are self-adjoint. By using this and the coercivity of A and C, (2.18) transforms into the estimate

$$c_0\|u(s)\|_H^2 + a_0\|\psi_s(0)\|_V^2 \leqslant \int_0^s \big\langle A'(t)\psi_s(t), \psi_s(t)\big\rangle + \big(C'(t)u(t), u(t)\big) + 2\big\langle B(t)u'(t) + Q(t)u(t), \psi_s(t)\big\rangle \, dt. \quad (2.19)$$

To be able to apply Grönwall's Lemma, we must only use $\|u(t)\|_H^2$ and $\|\psi_s(t)\|_V^2$ on the right-hand side of this inequality. Terms with A' and C' pose no difficulties, because

$$\int_0^s \big\langle A'(t)\psi_s(t), \psi_s(t)\big\rangle + \big(C'(t)u(t), u(t)\big) \, dt \leqslant \Lambda_1 \int_0^s \|\psi_s(t)\|_V^2 + \|u(t)\|_H^2 \, dt$$

with $\Lambda_1 = \|A'\|_{L^\infty(I;\mathcal{L}(V,V^*))} + \|C'\|_{L^\infty(I;\mathcal{L}(H))} > 0$. We must use differentiability of B to eliminate the dependence on $u'(t) = \psi_s''(t)$. Then, again using integration by parts (with vanishing boundary values $\psi_s(s) = \psi_s'(0) = 0$) we obtain

$$\begin{aligned}
\int_0^s \big(B(t)\psi_s''(t), \psi_s(t)\big) \, dt
&= -\int_0^s \big(B'(t)\psi_s'(t), \psi_s(t)\big) + \big(B(t)\psi_s'(t), \psi_s'(t)\big) \, dt \\
&\leqslant \Lambda_2 \int_0^s \|u(t)\|_H \|\psi_s(t)\|_H + \|u(t)\|_H^2 \, dt \\
&\leqslant \Lambda_2 \int_0^s \|\psi_s(t)\|_V^2 + 2\|u(t)\|_H^2 \, dt
\end{aligned}$$

with $\Lambda_2 > 0$. The treatment of Q with $Q(t) \in \mathcal{L}(H)$ is very easy because

$$\int_0^s (Q(t)u(t), \psi_s(t)) \, dt \leqslant \|Q\|_{L^\infty(I;\mathcal{L}(H))} \int_0^s \|u(t)\|_H^2 + \|\psi_s(t)\|_V^2 \, dt.$$

However, in the case $Q(t) \in \mathcal{L}(V,H)$, $Q(t)$ must not be applied to u, as this would create a dependence on $\|u(t)\|_V$. Integration by parts avoids this predicament, because

$$\begin{aligned}
&-\int_0^s \big(Q(t)\psi_s'(t), \psi_s(t)\big)\,dt \\
&= \big(Q(0)\psi_s(0), \psi_s(0)\big) + \int_0^s \big(Q'(t)\psi_s(t), \psi_s(t)\big) + \big(Q(t)\psi_s(t), \psi_s'(t)\big)\,dt \\
&\leqslant \Lambda_3\left(\|\psi_s(0)\|_V\|\psi_s(0)\|_H + \int_0^s \|\psi_s(t)\|_V^2 + \|u(t)\|_H^2\,dt\right)
\end{aligned}$$

holds, again with a constant $\Lambda_3 > 0$. We can proceed by noting that

$$\begin{aligned}
\|\psi_s(0)\|_V\|\psi_s(0)\|_H &\leqslant \varepsilon\|\psi_s(0)\|_V^2 + \varepsilon^{-1}\|\psi_s(0)\|_H^2 \\
&\leqslant \varepsilon\|\psi_s(0)\|_V^2 + \varepsilon^{-1}\left(\int_0^s \|u(t)\|_H\,dt\right)^2 \\
&\leqslant \varepsilon\|\psi_s(0)\|_V^2 + \varepsilon^{-1}T\int_0^s \|u(t)\|_H^2\,dt
\end{aligned}$$

holds for $\varepsilon > 0$. By choosing $\varepsilon = \Lambda_3^{-1}a_0/2$, this does not consume all of the coercivity on the left-hand side of (2.19).

We insert everything what we have learned into (2.19) to conclude that there exists $\Lambda_4 > 0$ that is independent of s such that

$$\|u(s)\|_H^2 + \|\psi_s(0)\|_V^2 \leqslant \Lambda_4 \int_0^s \|u(t)\|_H^2 + \|\psi_s(t)\|_V^2\,dt$$

holds in either case for the shape of Q. On the left we have $\psi_s(0)$, but on the right $\psi_s(t)$. We remedy this using $w(t) := \int_0^t u(\sigma)\,d\sigma$. Then $\psi_s(t) = w(t) - w(s)$ and $w(0) = 0$; in particular, $\psi_s(0) = -w(s)$. We obtain

$$\begin{aligned}
\|u(s)\|_H^2 + \|w(s)\|_V^2 &\leqslant \Lambda_4 \int_0^s \|u(t)\|_H^2 + \|w(t) - w(s)\|_V^2\,dt \\
&\leqslant 2s\Lambda_4\|w(s)\|_V^2 + 2\Lambda_4\int_0^s \|u(t)\|_H^2 + \|w(t)\|_V^2\,dt,
\end{aligned}$$

and after some reorganizing

$$\|u(s)\|_H^2 + (1 - 2s\Lambda_4)\|w(s)\|_V^2 \leqslant 2\Lambda_4\int_0^s \|u(t)\|_H^2 + \|w(t)\|_V^2\,dt.$$

Let $s_0 = (4\Lambda_4)^{-1}$, then $1 - 2s\Lambda_4 \geqslant 1/2$ for $s \in [0, s_0]$ and an application of Grönwall's Lemma shows $u(s) = 0$ for $s \in [0, s_0]$. After shifting the differential equation by $s_0/2$ we can show that the same estimates apply to $u(s_0/2 + \cdot)$, hence $u(s) = 0$ for $s \in [0, 3/2\, s_0]$. After a finite amount of such steps we arrive at $u(s) = 0$ in $[0, T]$. □

An additional consequence of the uniqueness is the fact that the whole sequence $(u_m)_{m\in\mathbb{N}}$ of Galerkin approximations has to converge weakly to u, not only a subsequence. Due to compactness of the embedding

$H^1(I;H) \cap L^2(I;V) \hookrightarrow L^2(I;H)$ one can even conclude that the u_m converge strongly to u, at least in $L^2(I;H)$.

We close this chapter by applying Theorem 2.10 to the wave equation that was introduced in Example 2.4 on page 11.

Example 2.11. Let $\Omega \subset \mathbb{R}^d$ be a domain. The problem

$$u''(t,x) - \operatorname{div}(a(t,x)\nabla u(t,x)) = f(t,x) \quad t \in (0,T),\ x \in \Omega$$
$$u = 0 \text{ on } (0,T) \times \partial\Omega,\ u(0,\cdot) = u_0,\ u'(0,\cdot) = u_1$$

possesses a unique weak solution $u \in L^2(I;H_0^1(\Omega)) \cap H^1(I;L^2(\Omega))$ for data $u_0 \in H_0^1(\Omega)$, $u_1 \in L^2(\Omega)$, $f \in L^2(I;L^2(\Omega))$, if the coefficient a belongs to $W^{1,\infty}(I;L^\infty(\Omega))$ and satisfies $a(t,x) \geqslant a_0 > 0$ for almost all $t \in I$ and $x \in \Omega$.

3 REGULARITY RESULTS

With the results we have given in Chapter 2, it is already possible to construct a well-defined solution operator that maps the operators A, B, C and Q to the solution u of (2.6), that is

$$(\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{A} + \mathcal{Q})u = f \quad \text{in } L^2(I; V^*), \tag{3.1a}$$
$$u(0) = u_0 \quad \text{and} \quad (\mathcal{C}u')(0) = u_1. \tag{3.1b}$$

However, as we will see in Chapter 4, they do not enable us to show Fréchet-differentiability of u with respect to the operators A and C. The lack of differentiability would severely limit the analysis of the inverse problems, and restrict the pool of applicable regularization algorithms. We give a short motivation what goes wrong for these parameters, and what we need to do in order to remedy the situation. An entirely formal differentiation of (3.1) with respect to A suggests that $u_h := (\partial_A u)[\bar{A}]$ might solve the equation

$$(\mathcal{C}u_h')' + \mathcal{B}u_h' + \mathcal{A}u_h + \mathcal{Q}u_h = -\bar{\mathcal{A}}u \tag{3.2}$$

with homogeneous initial values. Here, $\bar{\mathcal{A}}$ denotes the realization of the perturbation $\bar{A} \in W^{1,\infty}(I; \mathcal{L}(V, V^*))$ of A. The problem now lies in showing that this problem has a solution. The theory that we have developed in the previous chapter yields existence and uniqueness of u_h if the right-hand side either lies in $L^2(I; H)$ or in $H^1(I; V^*)$. However, since $u \in L^2(I; V)$ we can only infer $\bar{\mathcal{A}}u \in L^2(I; V^*)$. We have two options of dealing with this.

First, we can modify (3.2) in a way that makes it solvable for right-hand sides belonging to $L^2(I; V^*)$. We know that such right-hand sides are permissible for parabolic problems. By requiring $B(t)$ to be H-coercive we might cause the evolution equation to be "more parabolic". Indeed, this solves a lot of difficulties: The coercivity can be exploited to obtain $\|u_m'(t)\|_V$ on the left-hand side of the energy estimates for the Galerkin approximations (cf. equation (2.14)), hence no integration by parts is necessary for the treatment of $\langle f(t), u_m'(t)\rangle$ when $f(t) \in V^*$. In addition, this would cause the solution u to belong to $H^1(I; V)$, resulting in $\bar{\mathcal{A}}u \in H^1(I; V^*)$. Coercivity of B would therefore actually prevent the problem of occurring in two places at once. However, although being an easy fix, we feel that requiring B to be coercive restricts the class of applications too much. In particular, the acoustic wave equation (1.1) as proposed in the introduction would only satisfy this after we include an artificial dampening term $\nu u'$ with $\nu(t, x) \geqslant \nu_0 > 0$. Furthermore, this approach only solves the problems for A. If we linearize the equation with respect to C, then the right-hand side of (3.2) would read $(-\bar{\mathcal{C}}u')'$. We know that

$(\mathcal{C}u')' \in L^2(I;V^*)$, but have no similar results for arbitrary perturbations $\bar{C} \in W^{1,\infty}(I;\mathcal{L}(H))$.

The second way of attempting to deal with the problem is to enforce $\bar{\mathcal{A}}u \in H^1(I;V^*)$ and $(\bar{\mathcal{C}}u')' \in L^2(I;H)$ by showing that u is more regular than $L^2(I;V) \cap H^1(I;H)$. If $u \in H^1(I;V) \cap H^2(I;H)$, then u_h could be well-defined in both cases. We will also discover in Chapter 4 that even more regularity, that is $u \in H^2(I;V) \cap H^3(I;H)$, is necessary to show that u_h indeed characterizes the Fréchet-derivative of u. Hence, we should discuss suitable regularity results for the solution u of (3.1). Note that these are regularity results in *time*, because we neither need nor prove that $u(t)$ is smoother than V.

A basic result that one can easily obtain is the continuity of $u\colon I \to V$ and $u'\colon I \to H$, see for example [Sto01] or [DL00]. The special case $C = B = \mathrm{Id}$ is covered in [LM72]. Even for time-dependent operators this requires no additional smoothness of the operators, but the results clearly do not provide what we are looking for.

Beyond that, not much literature is available for the evolution equation (3.1). Kato [Kat88] shows regularity for systems $u'(t) + A(t)u(t) = f(t)$ using semigroup theory. Transforming (3.1a) to this shape would require (among other things) $C(t) \in \mathcal{L}(V)$. In specific PDEs this translates to *spatial* regularity of the corresponding coefficient, which then must be taken into account in the numerical inversion, thus unnecessarily complicating it.

The only source (to our knowledge) which actually deals with (3.1a) is the book [Lio61] by J. L. Lions. We present these results and how to obtain them in Section 3.1. In Section 3.2 we give a detailed discussion about the common method to show higher regularity in case of time-independent operators, why it does not work for our equation and what must be changed to remedy this. With these modifications, this approach is superior to the one discussed in the first subsection. Nevertheless, we decided to include Lions' results because the used techniques are uncommon and yield an elegant proof. Furthermore, the monograph [Lio61] is only available in French, and more importantly, only contains a proof for the parabolic case. Hence, our proof can even be considered as original work. Both regularity results can be used to solve the evolution equation with an additional operator acting on $(\mathcal{C}u')'$ as we will demonstrate in Section 3.3.

3.1 LIONS' LEMMA

In this subsection we turn to the regularity results that were stated by Lions [Lio61]. To avoid compatibility conditions, these results are restricted to homogeneous initial values $u_0 = u_1 = 0$ and homogeneous initial state of the right-hand side. To be more precise, his results are along the lines of the following theorem.

Theorem 3.1. *Let* $k \in \mathbb{N}$, $a_0, c_0 > 0$ *and suppose that the operators fulfill* $A \in W^{k+1,\infty}(I;\mathcal{L}^{sa}(V,V^*))$, $B \in W^{k,\infty}(I;\mathcal{L}(H))$, $C \in W^{k+1,\infty}(I;\mathcal{L}^{sa}(H))$

and $Q \in W^{k,\infty}(I;\mathcal{L}(V,H))$ *with* $A(t) \in \mathcal{L}^{sa}_{a_0}(V,V^*)$ *and* $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$ *for almost all* $t \in I$*. Further, let* $u_0 = u_1 = 0$ *and* $f \in H^k(I;H) \cup H^{k+1}(I;V^*)$ *with* $f(0) = \cdots = f^{(k-1)}(0) = 0$*. Then the unique solution* u *to problem* (3.1) *belongs to* $H^k(I;V) \cap H^{k+1}(I;H)$ *and satisfies* $u(0) = \cdots = u^{(k)}(0) = 0$*, as well as the estimate*

$$\|u\|^2_{H^k(I;V)} + \|u^{(k+1)}\|^2_{L^2(I;H)} \leqslant \Lambda \|f\|^2.$$

With $\|f\|$ *we denote either its* $H^k(I;H)$*- or its* $H^{k+1}(I;V^*)$*-norm. The constant* $\Lambda > 0$ *depends continuously on* a_0^{-1}, c_0^{-1} *and the operators* A*,* B*,* C*,* Q *in the spaces stated above.*

There are two main ingredients to Lions' approach: First, the extension of the time interval $I = [0,T]$ to $[0,\infty)$, which (together with the homogeneous initial values) allows integration by parts without being hindered by boundary terms. Second, it foregoes the Galerkin-approximation and instead solves the whole formulation directly using a variation of Lax-Milgram's lemma. Exponentially weighting in the function spaces ensures coercivity of the bilinear form. Thus, this technique is more reminiscent of the handling of elliptic PDEs than of hyperbolic equations.

We start by stating the required variant of Lax-Milgram.

Lemma 3.2 (Lions)**.** *Let* F *be a Hilbert space, and* $\Phi \subset F$ *a pre-Hilbert space that is continuously embedded in* F*. Let* $E\colon F \times \Phi \to \mathbb{R}$ *be a bilinear form that is continuous in the first variable, that is* $E(\cdot,\varphi) \in F^*$ *for all* $\varphi \in \Phi$*, and* $L \in \Phi^*$*. We further assume that there exists* $\alpha > 0$ *with* $E(\varphi,\varphi) \geqslant \alpha\|\varphi\|_\Phi$ *for all* $\varphi \in \Phi$*. Then there is at least one* $u \in F$ *that solves the equation*

$$E(u,\varphi) = L(\varphi), \qquad \varphi \in \Phi. \tag{3.3}$$

This solution u *satisfies the estimate* $\|u\|_F \leqslant (C_{\Phi\hookrightarrow F}/\alpha)\|L\|_{\Phi^*}$*, where* $C_{\Phi\hookrightarrow F}$ *denotes the norm of the embedding* $\Phi \hookrightarrow F$*.*

Proof. Can be found in [Lio61]. See also [Sho97] for a proof in english. □

Remarkable in this theorem, compared with the classical Lax-Milgram theorem, is that Φ does not have to be closed or dense in F and that $E(u,\cdot)$ does not have to be continuous for any u (although for the solution u of (3.3) it actually is). The downside is that we have to sacrifice the uniqueness of the solution. Fortunately, that is something we have already proven in Theorem 2.10 for our problem.

Let $k \in \mathbb{N}$ and $\gamma > 0$. The spaces in which we will apply Lions' lemma are

$$F := \left\{ u \in H^k_{loc}([0,\infty);V) \;\middle|\; e^{-\gamma t}u \in H^k_0([0,\infty);V) \cap H^{k+1}_0([0,\infty);H) \right\},$$
$$\Phi := \left\{ \varphi \in F \;\middle|\; e^{-\gamma t}\varphi \in H^{k+1}([0,\infty);V) \cap H^{k+2}([0,\infty);H) \right\}.$$

Here we abused notation, for example by writing $e^{-\gamma t}u$ instead of $e^{-\gamma \cdot}u$ or $t \mapsto e^{-\gamma t}u(t)$ in the hope that it increases readability. With $H^m_0([0,\infty);X)$ we denote the closure of $C^\infty_c([0,\infty);X)$ in the $H^m([0,\infty);X)$-norm. Hence, functions belonging to F have vanishing derivatives at zero up to order k. In particular, the extension $\tilde{u}\colon \mathbb{R} \to V$ of $u \in F$ by zero satisfies $e^{-\gamma t}\tilde{u} \in H^k([0,\infty);V) \cap H^{k+1}([0,\infty);H)$. In this setting, we can seek a solution u in F, but only have to test $E(u,\varphi) = L(\varphi)$ with functions φ that are more regular.

We equip F with the inner product

$$\begin{aligned}(u,v)_F \coloneqq \int_0^\infty &\Big(e^{-\gamma t}u^{(k)}(t), e^{-\gamma t}v^{(k)}(t)\Big)_V \\ &+ \Big(e^{-\gamma t}u^{(k+1)}(t), e^{-\gamma t}v^{(k+1)}(t)\Big)_H \,dt, \quad u,v \in F.\end{aligned}$$

The following lemma, which is similar to the Poincaré-inequality, shows that it is indeed sufficient to only include the derivatives of highest order.

Lemma 3.3. *Let X be a separable Hilbert space, $\gamma > 0$ and $m \in \mathbb{N}$. For every $\varphi \in H^m_{\mathrm{loc}}(\mathbb{R};X)$ with $e^{-\gamma t}\varphi^{(j)} \in L^2(\mathbb{R};X)$ $(j = 0,\dots,m)$ we have the estimate*

$$\big\|e^{-\gamma t}\varphi\big\|_{L^2(\mathbb{R};X)} \leqslant \gamma^{-m}\big\|e^{-\gamma t}\varphi^{(m)}\big\|_{L^2(\mathbb{R};X)}.$$

Proof. The set $C_c^\infty(\mathbb{R};X)$ is dense in the space of such φ if it is equipped with the norm $\|\varphi\| \coloneqq \|e^{-\gamma t}\varphi\|_{H^m(\mathbb{R};X)}$. Both sides of the asserted inequality are continuous in φ with respect to this norm, therefore we only have to show it for $\varphi \in C_c^\infty(\mathbb{R};X)$. Through integration by parts we see that

$$\int_{\mathbb{R}} e^{-2\gamma t}\big(\varphi'(t),\varphi(t)\big)_X\,dt = \gamma\int_{\mathbb{R}} e^{-2\gamma t}|\varphi(t)|^2\,dt.$$

The right-hand side is equal to $\gamma\big\|e^{-\gamma t}\varphi\big\|^2_{L^2(\mathbb{R};X)}$, and according to the Cauchy-Schwarz inequality the left-hand side is bounded from above by $\big\|e^{-\gamma t}\varphi'\big\|_{L^2(\mathbb{R};X)}\big\|e^{-\gamma t}\varphi\big\|_{L^2(\mathbb{R};X)}$. This proves the case $m = 1$; the assertion for $m > 1$ follows by induction. □

We regard Φ as a subspace of F by equipping it with the inner product of F. Note that Φ is not closed in F, thus $(\Phi,(\cdot,\cdot)_\Phi)$ is a pre-Hilbert space.

To construct E and L we first have to extend the operators and the right-hand side f from I to $[0,\infty)$. By extending $f \in H^k(I;H)$ to $t > T$ through its Taylor polynomial of order k at T, followed by a fast smooth decay to zero we can obtain an extension (which we will also denote by f) that fulfills

$$\|f\|_{H^k([0,\infty);H)} \leqslant 2\,\|f\|_{H^k(I;H)}.$$

Similar, in the case that $f \in H^{k+1}(I;V^*)$ we can expect the same inequality with $H^{k+1}([0,\infty);V^*)$ and $H^{k+1}(I;V^*)$ to hold. For the operators (and their derivatives) we only require L^∞-estimates in time, but we must ensure that $A(t)$ and $C(t)$ remain coercive for $t > T$. Therefore it is more advisable to smooth them to their values at T instead of zero. It is not necessary to perform this smoothing in technical detail, because it is obvious that we could do it in a way that the extension of A satisfies

$$\|A\|_{W^{k+1,\infty}([0,\infty);\mathcal{L}(V,V^*))} \leqslant 2\,\|A\|_{W^{k+1,\infty}(I;\mathcal{L}(V,V^*))},$$

analogous inequalities hold for B, C, Q and that the coercivity assumption

$$\langle A(t)v, v\rangle \geqslant \frac{a_0}{2}\|v\|_V^2, \quad (C(t)u, u) \geqslant \frac{c_0}{2}\|u\|_H^2, \quad t \in I, u \in H, v \in V$$

is fulfilled as well. On $F \times \Phi$ we define the bilinear form E as

$$\begin{aligned} E(u, \varphi) = &\int_0^\infty \left\langle e^{-\gamma t}(\mathcal{A}u + \mathcal{Q}u)^{(k)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right\rangle dt \\ &+ \int_0^\infty \left(e^{-\gamma t}(\mathcal{B}u')^{(k)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right) dt \\ &- \int_0^\infty \left((\mathcal{C}u')^{(k)}(t), \frac{d}{dt}\left(e^{-2\gamma t}\varphi^{(k+1)}(t)\right)\right) dt \end{aligned}$$

for $u \in F$, $\varphi \in \Phi$. It is continuous in the first component, but in general it will be discontinuous in the second component. For example, on first glance $E(u, \varphi)$ depends on the fact that $\varphi^{(k+1)}(t)$ belongs to V. This is incorporated in the space Φ but not in its norm.

If $f \in H^k(I; H)$, then the linear form L is given by

$$L(\varphi) = \int_0^\infty \left(e^{-\gamma t}f^{(k)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right) dt, \quad \varphi \in \Phi,$$

but in the case $f \in H^{k+1}(I; V^*)$ we must use

$$L(\varphi) = -\int_0^\infty \left\langle \frac{d}{dt}\left(e^{-2\gamma t}f^{(k)}(t)\right), \varphi^{(k)}(t)\right\rangle dt, \quad \varphi \in \Phi.$$

It follows easily that $L \in \Phi^*$ in both cases, with $\|L\|_{\Phi^*} \leqslant 2\|f\|_{H^k(I;H)}$ or $\|L\|_{\Phi^*} \leqslant 2(1 + 2\gamma)\|f\|_{H^{k+1}(I;V^*)}$. The only part that is missing in order to apply Lemma 3.2 is the coercivity of E, which we will verify next.

Lemma 3.4. *There is $\gamma_0 \geqslant 0$ that continuously depends on A, B, C, Q, a_0^{-1} and c_0^{-1} with the property that for $\gamma \geqslant \gamma_0$*

$$E(\varphi, \varphi) \geqslant \|\varphi\|_\Phi^2$$

holds for all $\varphi \in \Phi$.

Proof. In order to avoid tedious special cases we restrict our search to $\gamma \geqslant 1$. The term $E(\varphi, \varphi)$ naturally decomposes into four different parts, one for each operator. We estimate them individually.

(1) Repeated application of the product rule from Lemma 2.3 shows validity of the Leibniz rule for expressions with time-dependent linear operators. We employ it for the calculation of $(\mathcal{A}\varphi)^{(k)}$ in the integral that involves A to obtain

$$\begin{aligned} I_A &:= \int_0^\infty \left\langle e^{-\gamma t}(\mathcal{A}\varphi)^{(k)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right\rangle dt \\ &= \sum_{j=0}^{k} \binom{k}{j} \int_0^\infty \left\langle e^{-\gamma t}(\mathcal{A}^{(j)}\varphi^{(k-j)})(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right\rangle dt. \end{aligned}$$

The summand for $j = 0$ should yield coercivity, and in fact we have

$$\begin{aligned}
&\int_0^\infty \left\langle e^{-\gamma t}(\mathcal{A}\varphi^{(k)})(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right\rangle dt \\
&\quad = \frac{1}{2}\int_0^\infty e^{-2\gamma t}\frac{d}{dt}\left\langle A(t)\varphi^{(k)}(t), \varphi^{(k)}(t)\right\rangle \\
&\qquad\qquad - e^{-2\gamma t}\left\langle A'(t)\varphi^{(k)}(t), \varphi^{(k)}(t)\right\rangle dt \\
&\quad \geqslant \frac{1}{2}\left(\gamma a_0 - \|A'\|_{L^\infty([0,\infty);\mathcal{L}(V,V^*))}\right)\left\|e^{-\gamma t}\varphi^{(k)}\right\|^2_{L^2([0,\infty);V)}.
\end{aligned}$$

On the remaining terms we apply Lemma 3.3 to obtain

$$\begin{aligned}
&\int_0^\infty \left\langle e^{-\gamma t}A^{(j)}(t)\varphi^{(k-j)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right\rangle dt \\
&\quad = -\int_0^\infty \left\langle \frac{d}{dt}\left(e^{-2\gamma t}A^{(j)}(t)\varphi^{(k-j)}(t)\right), \varphi^{(k)}(t)\right\rangle dt \\
&\quad = -\int_0^\infty e^{-2\gamma t}\left\langle\left(A^{(j+1)}(t) - 2\gamma A^{(j)}(t)\right)\varphi^{(k-j)}(t), \varphi^{(k)}(t)\right\rangle \\
&\qquad\qquad + e^{-2\gamma t}\left\langle A^{(j)}(t)\varphi^{(k+1-j)}(t), \varphi^{(k)}(t)\right\rangle\ dt \\
&\quad \geqslant -2\|A\|_{W^{k+1,\infty}(I;\mathcal{L}(V,V^*))}(1 + 2\gamma + \gamma)\gamma^{-j}\left\|e^{-\gamma t}\varphi^{(k)}\right\|^2_{L^2([0,\infty);V)}
\end{aligned}$$

for $j = 1, \dots, k$. In this expression only the factor γ^{-j} still depends on j, and $(1 + 3\gamma)\gamma^{-j} \leqslant 4$ since $\gamma \geqslant 1$. Together with the summand for $j = 0$ we conclude

$$I_A \geqslant \left(\frac{\gamma a_0}{2} - 2^{k+4}\|A\|\right)\left\|e^{-\gamma t}\varphi^{(k)}\right\|^2_{L^2([0,\infty);V)}$$

with $\|A\| = \|A\|_{W^{k+1,\infty}(I;\mathcal{L}(V,V^*))}$.

(2) For Q we proceed in a similar way, but we may skip the special treatment of the zeroth summand and the integration by parts, because we can handle dependence on $\varphi^{(k+1)}$ in the H-norm. For $j = 0, \dots, k$ we deduce that

$$\begin{aligned}
&\int_0^\infty \left(e^{-\gamma t}Q^{(j)}(t)\varphi^{(k-j)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right) dt \\
&\quad \geqslant -2\|Q\|\gamma^{-j}\left\|e^{-\gamma t}\varphi^{(k)}\right\|_{L^2([0,\infty);V)}\left\|e^{-\gamma t}\varphi^{(k+1)}\right\|_{L^2([0,\infty);H)}
\end{aligned}$$

holds, where $\|Q\|$ denotes $\|Q\|_{W^{k,\infty}(I;\mathcal{L}(V,H))}$. Hence,

$$\begin{aligned}
I_Q &\coloneqq \int_0^\infty \left(e^{-\gamma t}(\mathcal{Q}\varphi)^{(k)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right) dt \\
&= \sum_{j=0}^k \binom{k}{j}\int_0^\infty \left(e^{-\gamma t}Q^{(j)}(t)\varphi^{(k-j)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\right) dt \\
&\geqslant -2^{k+1}\|Q\|\left\|e^{-\gamma t}\varphi^{(k)}\right\|_{L^2([0,\infty);V)}\left\|e^{-\gamma t}\varphi^{(k+1)}\right\|_{L^2([0,\infty);H)} \\
&\geqslant -2^k\|Q\|\left[\left\|e^{-\gamma t}\varphi^{(k)}\right\|^2_{L^2([0,\infty);V)} + \left\|e^{-\gamma t}\varphi^{(k+1)}\right\|^2_{L^2([0,\infty);H)}\right].
\end{aligned}$$

(3) The treatment of B is almost identical to that of Q, here we obtain

$$\begin{aligned} I_B &:= \int_0^\infty \Big(e^{-\gamma t}(\mathcal{B}\varphi')^{(k)}(t), e^{-\gamma t}\varphi^{(k+1)}(t)\Big)\,dt \\ &= \sum_{j=0}^{k}\binom{k}{j}\int_0^\infty \Big(e^{-\gamma t}(\mathcal{B}^{(j)}\varphi^{(k+1-j)})(t), e^{-\gamma t}\varphi^{(k+1)}(t)\Big)\,dt \\ &\geqslant -2^{k+1}\|B\|_{W^{k,\infty}(I;\mathcal{L}(H))}\Big\|e^{-\gamma t}\varphi^{(k+1)}\Big\|^2_{L^2([0,\infty);H)}. \end{aligned}$$

(4) For C we can again start with the product rule,

$$\begin{aligned} I_C &:= -\int_0^\infty \Big((\mathcal{C}\varphi')^{(k)}(t), \frac{d}{dt}\Big(e^{-2\gamma t}\varphi^{(k+1)}(t)\Big)\Big)\,dt \\ &= -\sum_{j=0}^{k}\binom{k}{j}\int_0^\infty \Big((\mathcal{C}^{(j)}\varphi^{(k+1-j)})(t), \frac{d}{dt}\Big(e^{-2\gamma t}\varphi^{(k+1)}(t)\Big)\Big)\,dt \end{aligned}$$

and separate the term for $j = 0$. Keeping in mind that $\varphi^{(k+1)}(0)$ does not have to vanish, we deduce

$$\begin{aligned} &-\int_0^\infty \Big((\mathcal{C}\varphi^{(k+1)})(t), \frac{d}{dt}\Big(e^{-2\gamma t}\varphi^{(k+1)}(t)\Big)\Big)\,dt \\ &\quad= \Big(C(0)\varphi^{(k+1)}(0), \varphi^{(k+1)}(0)\Big) \\ &\qquad+ \int_0^\infty \Big((\mathcal{C}\varphi^{(k+1)})'(t), e^{-2\gamma t}\varphi^{(k+1)}(t)\Big)\,dt \\ &\quad\geqslant c_0\Big\|\varphi^{(k+1)}(0)\Big\|_H^2 - 2\|C\|\Big\|e^{-\gamma t}\varphi^{(k+1)}\Big\|^2_{L^2([0,\infty);H)} \\ &\qquad+ \int_0^\infty e^{-2\gamma t}\Big(C(t)\varphi^{(k+2)}(t), \varphi^{(k+1)}(t)\Big)\,dt. \end{aligned}$$

Here we abbreviated the $W^{k+1,\infty}(I;\mathcal{L}(H))$-norm of C by $\|C\|$. The remaining integral can be dealt with by further integrating by parts. What we obtain is

$$\begin{aligned} &\int_0^\infty e^{-2\gamma t}\Big(C(t)\varphi^{(k+2)}(t), \varphi^{(k+1)}(t)\Big)\,dt \\ &\quad= \frac{1}{2}\int_0^\infty e^{-2\gamma t}\frac{d}{dt}\Big(C(t)\varphi^{(k+1)}(t), \varphi^{(k+1)}(t)\Big)\,dt \\ &\qquad- \frac{1}{2}\int_0^\infty e^{-2\gamma t}\Big(C'(t)\varphi^{(k+1)}(t), \varphi^{(k+1)}(t)\Big)\,dt \\ &\quad\geqslant -\frac{c_0}{2}\Big\|\varphi^{(k+1)}(0)\Big\|_H^2 + \Big(\frac{\gamma c_0}{2} - \|C\|\Big)\Big\|e^{-\gamma t}\varphi^{(k+1)}\Big\|^2_{L^2([0,\infty);H)}. \end{aligned}$$

In the other summands in I_C we do not have to consider boundary values and see that

$$\begin{aligned} &-\int_0^\infty \Big((\mathcal{C}^{(j)}\varphi^{(k+1-j)})(t), \frac{d}{dt}\Big(e^{-2\gamma t}\varphi^{(k+1)}(t)\Big)\Big)\,dt \\ &= \int_0^\infty e^{-2\gamma t}\Big((\mathcal{C}^{(j+1)}\varphi^{(k+1-j)} + \mathcal{C}^{(j)}\varphi^{(k+2-j)})(t), \varphi^{(k+1)}(t)\Big)\,dt \\ &\geqslant -2\|C\|(1+\gamma)\gamma^{-j}\Big\|e^{-\gamma t}\varphi^{(k+1)}\Big\|^2_{L^2([0,\infty);H)} \end{aligned}$$

holds for $j = 1, \dots, k$. Combining the above calculations and using $(1+\gamma)\gamma^{-j} \leqslant 2$ for $j \geqslant 1$ we close with

$$I_C \geqslant \left(\frac{\gamma c_0}{2} - 2^{k+2}\|C\|\right) \left\|e^{-\gamma t}\varphi^{(k+1)}\right\|^2_{L^2([0,\infty);H)}.$$

Finally, we put everything together to see that

$$\begin{aligned} E(\varphi, \varphi) &= I_A + I_B + I_C + I_Q \\ &\geqslant \left(\frac{\gamma a_0}{2} - 2^k\left[16\|A\| - \|Q\|\right]\right) \left\|e^{-\gamma t}\varphi^{(k)}\right\|^2_{L^2([0,\infty);V)} \\ &\quad + \left(\frac{\gamma c_0}{2} - 2^k\left[4\|C\| - 2\|B\| - \|Q\|\right]\right) \left\|e^{-\gamma t}\varphi^{(k+1)}\right\|^2_{L^2([0,\infty);H)}. \end{aligned}$$

The assertion follows by taking $\gamma_0 \geqslant 1$ and large enough to force both factors to be greater than one. $\square$

At last, we can prove the regularity theorem that was stated on page 25.

Proof of Theorem 3.1. We define γ to be the constant γ_0 from Lemma 3.4 and apply Lemma 3.2 in order to obtain $u \in F$ that solves

$$E(u, \varphi) = L(\varphi), \quad \varphi \in \Phi, \tag{3.4}$$

and claim that the restriction $\tilde{u} := u|_{[0,T]}$ is a solution of (3.1) that satisfies the asserted estimate. The latter is easy to verify. According to Lemma 3.2, the particular solution u fulfills $\|u\|_F \leqslant \|L\|_{\Phi^*}$. Thus, $\|u\|_F \leqslant 6\gamma_0\|f\|$ (in both possibilities for f). Furthermore,

$$\begin{aligned} \|\tilde{u}^{(k+1)}\|^2_{L^2(I;H)} &= e^{2\gamma_0 T} \int_0^T \left\|e^{-\gamma_0 T} u^{(k+1)}(t)\right\|^2_H dt \\ &\leqslant e^{2\gamma_0 T} \int_0^\infty \left\|e^{-\gamma_0 t} u^{(k+1)}(t)\right\|^2_H dt \end{aligned}$$

and due to Lemma 3.3 we have for $j = 0, \dots, k$

$$\|\tilde{u}^{(j)}\|^2_{L^2(I;V)} \leqslant e^{2\gamma_0 T}\gamma_0^{2(j-k)} \int_0^\infty \left\|e^{-\gamma_0 t} u^{(k)}(t)\right\|^2_V dt.$$

In this way we obtain the estimate

$$\|\tilde{u}\|^2_{H^k(I;V)} + \|\tilde{u}^{(k+1)}\|^2_{L^2(I;H)} \leqslant e^{2\gamma_0 T}\|u\|_F^2 \leqslant 36\gamma_0^2 e^{2\gamma_0 T}\|f\|^2,$$

in which the factor in front of $\|f\|^2$ continuously depends on the operators and the inverses of the coercivity constants.

The function u and its derivatives up to order k vanish at zero due to the design of the space F. Hence, the asserted initial conditions are satisfied by $\tilde{u}$. To verify that the weak formulation holds, let $v \in V$ and $\vartheta \in C_c^\infty(I)$. We extend ϑ to $[0,\infty)$ by zero and again denote the resulting function by ϑ. Furthermore, we define $Y(t) := t^k/k!$ if $t \geqslant 0$ and $Y(t) := 0$ otherwise to construct $\psi := Y * \vartheta$. Clearly, $e^{-\gamma t}\psi^{(j)} \in L^2([0,\infty))$ for all $0 \leqslant j \leqslant k$, $\psi(0) = \cdots = \psi^{(k)}(0) = 0$ and $\psi^{(k+1)} = \vartheta$, because

$\psi^{(k)}(s) = \int_0^s \vartheta(t)\,dt$. In particular, via $\varphi(t) := \psi(t)v$ we obtain $\varphi \in \Phi$ which we can plug into (3.4). The result, after integrating by parts if $f \in H^{k+1}(I; V^*)$, is

$$\begin{aligned}&\int_0^\infty \left\langle f^{(k)}(t), v\right\rangle e^{-2\gamma t}\vartheta(t)\,dt \\ &= \int_0^\infty \left\langle (\mathcal{A}u)^{(k)}(t), v\right\rangle e^{-2\gamma t}\vartheta(t) + \left((\mathcal{Q}u)^{(k)}(t), v\right) e^{-2\gamma t}\vartheta(t)\,dt \\ &\quad + \int_0^\infty \left((\mathcal{B}u')^{(k)}(t), v\right) e^{-2\gamma t}\vartheta(t) - \left((\mathcal{C}u')^{(k)}(t), v\right) \frac{d}{dt}\left(e^{-2\gamma t}\vartheta(t)\right) dt.\end{aligned}$$

Since $\vartheta \in C_c^\infty(I)$ is arbitrary, it follows that

$$\frac{d^k}{dt^k}\left(\left\langle (\mathcal{C}\tilde{u}')'(t) + B(t)\tilde{u}'(t) + A(t)\tilde{u}(t) + Q(t)\tilde{u}(t) - f(t), v\right\rangle\right) = 0$$

holds in I in the distributional sense. Integrating this k times shows that $\tilde{u}$ fulfills the evolution equation. □

3.2 DIFFERENTIATING THE EQUATION WITH RESPECT TO TIME

A prominent approach to obtain higher regularity is the formal differentiation of both sides of the evolution equation with respect to time. For time-independent operators, this approach is for example used in [Bré11; KR14a; Wlo87]. In order to ease the presentation of the ideas, let us first consider the evolution equation $u'' + \mathcal{A}u = f$. By differentiating in time we obtain

$$u'''(t) + A'(t)u(t) + A(t)u'(t) = f'(t),$$

which can then be treated as a new problem for $v := u'$. The expression $A'(t)u(t)$ is regarded as independent of v and moved to the right-hand side. After furnishing the equation with suitable initial values, the resulting problem for v reads

$$v'' + \mathcal{A}v = f' - \mathcal{A}'u \quad \text{in } L^2(I; V^*) \tag{3.5a}$$

$$v(0) = u_1,\ v'(0) = f(0) - A(0)u_0 \tag{3.5b}$$

The idea is to show that such a v exists and that is has to be equal to u'. However, to ensure the existence of such a v, we know from the previous sections that we need $t \mapsto A'(t)u(t)$ to be an element of either $L^2(I; H)$ or $H^1(I; V^*)$. The latter would require $u \in H^1(I; V)$, which is what we are trying to obtain through these regularity considerations. The former can be fulfilled by assuming $A'(t) \in \mathcal{L}(V, H)$. Broadly speaking, in applications this forces coefficients that influence second-order spatial derivatives to be *independent* of time. In the case of the wave equation, this would imply that $A(t)u = \Delta u + q(t)u$ is permitted, while $A(t)u = \operatorname{div}(\nabla u/\rho(t))$ is not. Unfortunately, the latter case is very interesting for us since it involves a time-dependent mass density.

The precise result that is achieved by this approach is shown in the following theorem.

Theorem 3.5. *Let $a_0 > 0$ and suppose that $A \in W^{1,\infty}(I; \mathcal{L}^{sa}(V, V^*))$ with $A' \in L^\infty(I; \mathcal{L}(V, H))$ and $A(t) \in \mathcal{L}^{sa}_{a_0}(V, V^*)$ for almost all $t \in I$. Further, let $u_0 \in V$, $u_1 \in V$ and $f \in H^1(I; H) \cup H^2(I; V^*)$ with $u_2 := f(0) - A(0)u_0 \in H$. Then the solution u of*

$$u'' + \mathcal{A}u = f \quad \text{in } L^2(I; V^*) \tag{3.6a}$$

$$u(0) = u_0,\ u'(0) = u_1 \tag{3.6b}$$

belongs to $W^{1,\infty}(I; V) \cap W^{2,\infty}(I; H)$ and fulfills

$$\|u\|^2_{W^{1,\infty}(I;V)} + \|u''\|^2_{L^\infty(I;H)} \leqslant \Lambda \left(\|u_0\|^2_V + \|u_1\|^2_V + \|u_2\|^2_H + \|f\|^2 \right),$$

with the constant $\Lambda > 0$ continuously depending on a_0^{-1} and A.

Proof. The assumptions on the operators and the data enable us to use Theorem 2.10 to obtain unique solutions $u, v \in L^2(I; V) \cap H^1(I; H)$ of problems (3.6) and (3.5), respectively. As already hinted at, we define $w(t) := u_0 + \int_0^t v(s)\,ds$ and proceed to show that w must equal u. The initial conditions $w(0) = u_0$ and $w'(0) = v(0) = u_1$ are easily verified. Furthermore, for almost all $t \in I$ we calculate

$$\begin{aligned} w''(t) &= v'(0) + \int_0^t v''(s)\,ds \\ &= f(0) - A(0)u_0 + \int_0^t f'(s) - A'(s)u(s) - A(s)v(s)\,ds \\ &= f(t) - (\mathcal{A}w)(0) - \int_0^t A'(s)u(s) + (\mathcal{A}w)'(s) - A'(s)w(s)\,ds \\ &= f(t) - A(t)w(t) - \int_0^t A'(s)(u(s) - w(s))\,ds, \end{aligned}$$

where the equalities hold in V^*. Hence, the difference $h := w - u$ of w and u solves $h''(t) + A(t)h(t) = \int_0^t A'(s)h(s)\,ds$ and possesses homogeneous initial values. Let $T^* \in I = [0, T]$ be maximal with the property that $h = 0$ almost everywhere on $[0, T^*]$. We prove $u = w$ by showing that $T^* = T$. The asserted estimate for u is then a direct consequence of the estimates for u and $u' = v$ obtained by the application of Theorem 2.10.

Suppose that $T^* < T$ and let $T_0 \in (T^*, T)$. According to the energy estimates for h on the subinterval $[0, T_0] \subset I$, there exists $\Lambda > 0$ (depending on T, not on T_0) such that

$$\begin{aligned} \int_{T^*}^{T_0} \|h(t)\|^2_V\,dt &\leqslant \Lambda \int_{T^*}^{T_0} \left(\int_{T^*}^{s} \|A'(t)h(t)\|_H\,dt \right)^2 ds \\ &\leqslant \Lambda \|A'\|^2_{L^\infty(I;\mathcal{L}(V,H))} \int_{T^*}^{T_0} (s - T^*) \int_{T^*}^{s} \|h(t)\|^2_V\,dt\,ds \\ &= \frac{\Lambda}{2} \|A'\|^2_{L^\infty(I;\mathcal{L}(V,H))} (T_0 - T^*)^2 \int_{T^*}^{T_0} \|h(t)\|^2_V\,dt. \end{aligned}$$

By design of T^*, the term $\int_{T^*}^{T_0} \|h(t)\|^2_V\,dt$ is positive, therefore we have $2 \leqslant \Lambda \|A'\|^2 (T_0 - T^*)^2$ for all $T_0 \in (T^*, T)$. Observing that the right-hand side tends to zero as $T_0 \to T^*$ concludes the proof. $\square$

The generalization of this regularity result to nontrivial operators $B \in W^{1,\infty}(I;\mathcal{L}(H))$, $C \in W^{2,\infty}(I;\mathcal{L}(H))$ and $Q \in W^{1,\infty}(I;\mathcal{L}(V,H))$ is straightforward.

Together with S. Grützner we proposed an improvement of this theorem that allows $A'(t) \in \mathcal{L}(V,V^*)$ in [GG19]. The rest of this section is devoted to the approach we used there. The main idea is to note that it seems like an unnecessary detour to derive an equation for $v = u'$, but while solving it seeing u and v as unrelated, only to then having to prove $u' = v$. In particular, if we replace $u(t)$ by $u_0 + \int_0^t v(s)\,ds$ in the evolution equation (3.5a) for v, then the term $\mathcal{A}'u$ would not pose a problem since we expect v to belong to $L^2(I;V)$. The result is a mixed integral and differential equation for v, which reads

$$v''(t) + A(t)v(t) + \int_0^t A'(t)v(s)\,ds = f'(t) - A'(t)u_0 \tag{3.7}$$

for $t \in I$. The dependence on u has disappeared in this equation, but we can recover it by showing that $u_0 + \int_0^t v(s)\,ds$ solves the problem (3.6) that is already uniquely solved by u. This concept seems simple, but we were unable to find any regularity results that build on this idea in literature.

The main disadvantage of this approach is a more involved proof: Existence of v does not immediately follow from the theory we developed in Section 2.2. Both the Galerkin equations and the energy estimates have to be amended to work with an additional integral term. Furthermore, even higher regularity for u does not follow immediately because the regularity result itself cannot be applied to (3.7) due to an additional term with $\int_0^t \int_0^s v(\sigma)\,d\sigma\,ds$ that appears after another differentiation in time. To be able to obtain a result $u \in H^k(I;V) \cap H^{k+1}(I;H)$ (similar to Theorem 3.1) with $k \in \mathbb{N}$, we have to support evolution equations with integral terms up to order k. To denote these integral terms we introduce the operator

$$R_{X,g}\colon L^2(I;X) \to H^1(I;X), \quad (R_{X,g}v)(t) := g + \int_0^t v(s)\,ds, \quad t \in I$$

where X is a separable Banach space and $g \in X$. We will make frequent use of $R_{V,g}$ and thus choose to abbreviate it as R_g.

The operators B, C and Q also generate time-dependent zero order terms after differentiating the equation more than once, but since these operators map to H, they can safely be moved to the right-hand side of the equation. Thus, they do not change the technique, only complicate the presentation and the compatibility conditions. However, the inverse problems that we will encounter will not affect exactly one of the four operators (the mass density ρ in Chapter 5 will influence three of them). Hence, we must include them in further considerations.

We begin by establishing the existence of a solution v to the evolution problem

$$(\mathcal{C}v')' + \mathcal{B}v' + (\mathcal{A}+\mathcal{Q})v + \sum_{i=1}^{k}(\mathcal{D}_i + \mathcal{E}_i)(R_0^i v) = f \text{ in } L^2(I;V^*), \tag{3.8a}$$

$$v(0) = v_0 \quad \text{and} \quad (\mathcal{C}v')(0) = v_1. \tag{3.8b}$$

with $k \in \mathbb{N}$, $D_i(t) \in \mathcal{L}(V,V^*)$ and $E_i(t) \in \mathcal{L}(V,H)$ for $i = 1,\dots,k$ and almost all $t \in I$. We will try to keep the presentation short by only indicating how the corresponding results and proofs from Section 2.2 have to be modified. A self-contained and more detailed analysis can be found in [GG19]. Let $(\varphi_i)_{i\in\mathbb{N}} \subset V$ again be an orthonormal basis of H and $V_m := \operatorname{lin}\{\,\varphi_j \mid j = 1,\dots,m\,\}$. Instead of Theorem 2.5 we obtain the following result about the existence of approximate solutions in V_m.

Theorem 3.6. *Let $m \in \mathbb{N}$, $c_0 > 0$ and suppose that $A \in L^\infty(I;\mathcal{L}(V,V^*))$, $B \in L^\infty(I;\mathcal{L}(H))$, $Q \in L^\infty(I;\mathcal{L}(V,H))$ and $C \in W^{1,\infty}(I;\mathcal{L}^{sa}(H))$. Further, let $v_0, v_1 \in H$, $f \in L^2(I;V^*)$ and $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$ for almost all $t \in I$. The new operators have to fulfill $D_i \in L^\infty(I;\mathcal{L}(V,V^*))$ and $E_i \in L^\infty(I;\mathcal{L}(V,H))$ for $i = 1,\dots,k$. Then there exists a unique solution $v_m \in H^2(I;V_m)$ of the equation*

$$(\mathcal{C}v_m')' + \mathcal{B}v_m' + (\mathcal{A}+\mathcal{Q})v_m + \sum_{i=1}^{k}(\mathcal{D}_i + \mathcal{E}_i)(R_0^i v_m) = f \quad \textit{in } L^2(I;V_m^*) \tag{3.9a}$$

that also satisfies the initial conditions

$$v_m(0) = \sum_{j=1}^{m}(v_0,\varphi_j)\varphi_j \quad \textit{and} \quad v_m'(0) = \sum_{j=1}^{m}\Big(C(0)^{-1}v_1,\varphi_j\Big)\varphi_j. \tag{3.9b}$$

Proof. In the proof of Theorem 2.5, we derived the system (2.8) of m second-order ordinary differential equations for the coefficients $\alpha = (\alpha_j)_{j=1,\dots,m}$ of $v_m(t) = \sum_{j=1}^m \alpha_j(t)\varphi_j$, together with initial conditions for α and α'. Due to the dependence on $R_0^l v_m$, we introduce $l \cdot m$ additional unknowns $\gamma^{[l]} \in \mathbb{R}^m$ for $l = 1,\dots,k$. They are linked to each other and α by the $l \cdot m$ first-order equations

$$(\gamma^{[1]})'(t) = \alpha(t), \quad (\gamma^{[l]})'(t) = \gamma^{[l-1]}(t),\ l = 2,\dots,k$$

and have to satisfy homogeneous initial conditions $\gamma^{[l]}(0) = 0 \in \mathbb{R}^m$ for $l = 1,\dots,k$. These ensure that $\gamma^{[l]} = R^l_{0,\mathbb{R}^m}\alpha$. The initial conditions for α and α' remain unchanged, but we supplement the second-order system (2.8) to read

$$M_2(t)\cdot\alpha''(t) + M_1(t)\cdot\alpha'(t) + M_0(t)\cdot\alpha(t) + \sum_{l=1}^{k} M_{-l}(t)\cdot\gamma^{[l]}(t) = F(t).$$

The new matrices $M_{-l}(t) \in \mathbb{R}^{m\times m}$ $(l = 1,\dots,k)$ are defined as

$$(M_{-l}(t))_{ij} = \big\langle (D_l(t) + E_l(t))\varphi_j, \varphi_i\big\rangle, \quad i,j = 1,\dots,m.$$

By the same arguments that we made for Theorem 2.5, there is a unique solution $\alpha \in H^2(I;\mathbb{R}^m)$, and due to the design of the $\gamma^{[l]}$ it easily follows that $v_m(t) = \sum_{j=1}^m \alpha_j(t)\varphi_j$ solves the Galerkin problem (3.9). □

As for the original evolution equation, establishing an upper bound for v_m (independent of $m \in \mathbb{N}$) enables us to let m tend to infinity to obtain a solution v.

Lemma 3.7. *Let* $a_0, c_0 > 0$ *and* $A \in W^{1,\infty}(I;\mathcal{L}^{sa}(V,V^*))$, $B \in L^\infty(I;\mathcal{L}(H))$, $C \in W^{1,\infty}(I;\mathcal{L}^{sa}(H))$, $Q \in L^\infty(I;\mathcal{L}(V,H))$. *Furthermore, we expect* $D_i \in W^{1,\infty}(I;\mathcal{L}(V,V^*))$ *and* $E_i \in L^\infty(I;\mathcal{L}(V,H))$ *for* $i = 1,\dots,k$. *Lastly, let* $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$, $A(t) \in \mathcal{L}^{sa}_{a_0}(V,V^*)$ *for almost all* $t \in I$, $f \in L^2(I;H) \cup H^1(I;V^*)$, $v_0 \in V$ *and* $v_1 \in H$. *Under these assumptions problem (3.8) has a solution* $v \in L^2(I;V) \cap H^1(I;H)$ *that satisfies the energy estimate*

$$\operatorname*{ess\,sup}_{t\in I} \|v(t)\|_V^2 + \|v'(t)\|_H^2 \leqslant \Lambda \left(\|v_0\|_V^2 + \|v_1\|_H^2 + \|f\|^2 \right),$$

where $\Lambda > 0$ *continuously depends on* c_0^{-1}, a_0^{-1} *and the operators.*

Proof. We take $v_m \in H^2(I;V_m)$ to be the unique solution to the Galerkin equations (3.9) and derive the energy estimate for them by testing the equation with $v_m'(t) \in V_m$. The difference to the proof of Theorem 2.8 are the extra terms

$$-\sum_{i=1}^k \int_0^t \langle D_i(s)(R_0^i v_m)(s), v_m'(s)\rangle + \big(E_i(s)(R_0^i v_m)(s), v_m'(s)\big)\, ds$$

that appear on the right-hand side of (2.13) on page 15. We modify them in order to again obtain the integral inequality (2.14) that yield the estimates by invoking Grönwall's lemma. But first let us note that for $i \geqslant 2$, $s \in I$ and $g \in L^2(I;V)$

$$\|(R_0^i g)(s)\| \leqslant (R_{0,\mathbb{R}}^i \|g\|)(s) = \int_0^s (R_{0,\mathbb{R}}^{i-1}\|g\|)(\sigma)\, d\sigma \leqslant s(R_{0,\mathbb{R}}^{i-1}\|g\|)(s)$$

holds because $s \mapsto (R_{0,\mathbb{R}}^{i-1}\|g\|)(s)$ is increasing. In particular, we see that

$$\begin{aligned}\int_0^t \|(R_0^i v_m)(s)\|_V \|v_m(s)\|_V\, ds &\leqslant T^{i-1} \int_0^t \|v_m(s)\|_V \int_0^s \|v_m(\sigma)\|_V\, d\sigma\, ds \\ &\leqslant T^i \int_0^t \|v_m(s)\|_V^2\, ds\end{aligned}$$

is true for $t \in I$ and $i \geqslant 1$. For the terms that include D_i we use this formula and integrate by parts to eliminate the dependence on the V-norm of $v_m'(s)$. This yields

$$\begin{aligned}
-\int_0^t \langle D_i(s)(R_0^i v_m)(s), v_m'(s)\rangle \,ds &= -\langle D_i(t)(R_0^i v_m)(t), v_m(t)\rangle \\
&\quad + \int_0^t \langle D_i'(s)(R_0^i v_m)(s), v_m(s)\rangle - \langle D_i(s)(R_0^{i-1} v_m)(s), v_m(s)\rangle \,ds \\
&\leqslant \|D_i\| \|(R_0^i v_m)(t)\|_V \|v_m(t)\|_V \\
&\quad + \|D_i\| \int_0^t \big(\|(R_0^i v_m)(s)\|_V + \|(R_0^{i-1} v_m)(s)\|_V\big) \|v_m(s)\|_V \,ds \\
&\leqslant T^{i-1} \|D_i\| \left(\|v_m(t)\|_V \int_0^t \|v_m(s)\|_V \,ds + (T+1) \int_0^t \|v_m(s)\|_V^2 \,ds \right) \\
&\leqslant T^{i-1} \|D_i\| \left(\varepsilon \|v_m(t)\|_V^2 + (T + T/\varepsilon + 1) \int_0^t \|v_m(s)\|_V^2 \,ds \right)
\end{aligned}$$

for arbitrary $\varepsilon > 0$ and $i = 1, \dots, k$ when D_i is measured with the norm of $W^{1,\infty}(I; \mathcal{L}(V, V^*))$. These terms can be distributed to the left- and right-hand sides of (2.14) without difficulty. The fact that $E_i(t) \in \mathcal{L}(V, H)$ allows it to be treated similar to Q, that is

$$\begin{aligned}
-2 \int_0^t \langle E_i(s)(R_0^i v_m)(s), v_m'(s)\rangle \,ds \\
\leqslant \|E_i\|_{L^\infty(I;\mathcal{L}(V,H))} T^i \int_0^t \|v_m(s)\|_V^2 + \|v_m'(s)\|_H^2 \,ds,
\end{aligned}$$

which poses no problems when added to the right-hand side of (2.14), thus the estimates for v_m still hold.

Due to these energy estimates, the sequence $(v_m)_{m\in\mathbb{N}}$ must possess a weakly convergent subsequence. The arguments that prove this weak limit to be a solution of (3.8) are exactly the same as in the proof of Theorem 2.9. □

For our purposes we do not require uniqueness of this solution, and would not be able to show it under the assumptions of the theorem above because of missing differentiability of B and Q in time.

With these preliminaries out of the way, we can go back to our original evolution problem (3.1), that is

$$\begin{aligned}
(\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{A} + \mathcal{Q})u &= f \quad \text{in } L^2(I; V^*), \\
u(0) = u_0 \ \text{ and } \ (\mathcal{C}u')(0) &= u_1.
\end{aligned}$$

Formally differentiating the equation k times with respect to time and sorting the resulting terms leads to

$$\begin{aligned}
f^{(k)} &= (\mathcal{C}u^{(k+1)})' + (k\mathcal{C}' + \mathcal{B})u^{(k+1)} \\
&\quad + \left(\mathcal{A} + \mathcal{Q} + k\mathcal{B}' + \frac{k(k+1)}{2}\mathcal{C}'' \right) u^{(k)} \\
&\quad + \sum_{j=1}^{k} \left[\binom{k}{j} (\mathcal{A}^{(j)} + \mathcal{Q}^{(j)}) + \binom{k}{j+1} \mathcal{B}^{(j+1)} + \binom{k+1}{j+2} \mathcal{C}^{(j+2)} \right] u^{(k-j)}
\end{aligned} \tag{3.10}$$

as an equality in $L^2(I;V^*)$. Note that the dependence on $C^{(k+2)}$ and $B^{(k+1)}$ is merely notational (to avoid multiple summations with different ranges) because the coefficients in front of them vanish. This equation can also be exploited to see that $u^{(k)}$ might possess the initial values

$$u^{(k)}(0) = u_k, \quad (\mathcal{C}u^{(k+1)})(0) = C(0)u_{k+1},$$

which are given recursively for $k \geqslant 0$ via

$$\begin{aligned} C(0)u_{k+2} = f^{(k)}(0) - ((k+1)C'(0) + B(0))u_{k+1} \\ - \sum_{j=0}^{k} \Bigg[\binom{k}{j} (A^{(j)}(0) + Q^{(j)}(0)) \\ + \binom{k}{j+1} B^{(j+1)}(0) + \binom{k+1}{j+2} C^{(j+2)}(0) \Bigg] u_{k-j}, \end{aligned} \tag{3.11}$$

starting from the already given u_0 and u_1. Since $C \in W^{1,\infty}(I;\mathcal{L}^{sa}(H))$ it is continuous, which makes $C(0) \in \mathcal{L}^{sa}(H)$ well-defined. Moreover, from the coercivity of $C(t)$ we are able to conclude that $C(0)$ is invertible. Therefore u_{k+2} is well-defined as long as the right-hand side of the above equation is an element of H. This requirement is not easily fulfilled, because even for $u_j \in V$ $(j = 0, \dots, k+1)$, the right-hand side (in general) only belongs to V^*. Requiring that equation (3.11) is solvable is therefore a compatibility condition for the initial values u_0, u_1 and the values of both the right-hand side f and the operators (including their derivatives) at the initial time.

We interpret (3.10) as a second-order evolution equation for $v := u^{(k)}$ without moving lower order derivatives of u to the right-hand side. Instead, they are replaced by consecutive applications of R_{u_l} $(l = m, \dots, n)$ to v, which we write as

$$\bigcirc_{l=m}^{n} R_{u_l} v := \begin{cases} (R_{u_m} \circ \cdots \circ R_{u_n}) v & \text{if } m \leqslant n, \\ v & \text{if } m > n. \end{cases}$$

Hence, we have the auxiliary equation

$$\begin{aligned} f^{(k)} = (\mathcal{C}v')' + (k\mathcal{C}' + \mathcal{B})v' + (\mathcal{A} + \mathcal{Q})v + \left(k\mathcal{B}' + \frac{k(k+1)}{2}\mathcal{C}'' \right) v \\ + \sum_{j=1}^{k} \Bigg[\binom{k}{j} (\mathcal{A}^{(j)} + \mathcal{Q}^{(j)}) + \binom{k}{j+1} \mathcal{B}^{(j+1)} \\ + \binom{k+1}{j+2} \mathcal{C}^{(j+2)} \Bigg] \bigcirc_{l=k-j}^{k-1} R_{u_l} v \end{aligned} \tag{3.12a}$$

and initial values

$$v(0) = u_k \text{ in } H, \; (\mathcal{C}v')(0) = C(0)u_{k+1} \text{ in } V^*. \tag{3.12b}$$

Existence of such a v is now an almost immediate consequence of Lemma 3.7.

Corollary 3.8. *Let $k \in \mathbb{N}$, $a_0, c_0 > 0$. Suppose $A \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*))$, $B \in W^{k,\infty}(I; \mathcal{L}(H))$, $C \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(H))$ and $Q \in W^{k,\infty}(I; \mathcal{L}(V, H))$ with $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$, $A(t) \in \mathcal{L}^{sa}_{a_0}(V, V^*)$ for almost all $t \in I$. Furthermore, we assume $f \in H^k(I; H) \cup H^{k+1}(I; V^*)$, $u_j \in V$ for $j = 0, \dots, k$ and $u_{k+1} \in H$. Then there exists $v \in L^2(I; V) \cap H^1(I; H)$ that solves (3.12) and satisfies*

$$\|v\|^2_{L^\infty(I;V)} + \|v'\|^2_{L^\infty(I;H)} \leqslant \Lambda \left(\sum_{j=0}^{k} \|u_j\|^2_V + \|u_{k+1}\|^2_H + \|f\|^2 \right) \tag{3.13}$$

where $\Lambda > 0$ depends continuously on c_0^{-1}, a_0^{-1} and the operators A, B, C, Q.

Proof. To obtain (3.8) from (3.12), we set

$$\begin{aligned} \tilde{B} &= B + C', \qquad \tilde{Q} = Q + kB' + \frac{k(k+1)}{2} C'', \\ D_j &= \binom{k}{j} A^{(j)}, \quad E_j = \binom{k+1}{j+2} C^{(j+2)} + \binom{k}{j+1} B^{(j+1)} + \binom{k}{j} Q^{(j)}, \end{aligned}$$

for $j = 1, \dots, k$. The operators A and C are left as-is and the initial values are $v_0 = u_k$ and $v_1 = u_{k+1}$. The right-hand side of (3.12a) is not yet linear in v due to inhomogeneities that stem from R_{u_l}. By writing

$$\begin{aligned} \left(\bigcirc_{l=k-j}^{k-1} R_{u_l} v \right)(t) &= (R_0^j v)(t) + \sum_{l=k-j}^{k-1} (R_0^{k-1-l} u_l)(t) \\ &= (R_0^j v)(t) + \sum_{l=k-j}^{k-1} u_l \frac{t^{k-1-l}}{(k-1-l)!} \end{aligned}$$

they can be moved to the left-hand side of the equation. Thus, it consists of

$$\tilde{f}(t) = f^{(k)}(t) - \sum_{j=1}^{k} \left[(D_j(t) + E_j(t)) \sum_{l=k-j}^{k-1} u_l \frac{t^{k-1-l}}{(k-1-l)!} \right].$$

The assertion follows by utilizing Lemma 3.7. □

We still have to show that v is equal to $u^{(k)}$. The next lemma permits to show this through induction over k.

Lemma 3.9. *Suppose v is a solution to problem (3.12) with $k \in \mathbb{N}$, and that $u_{k-1} \in V$. Then $R_{u_{k-1}} v$ solves (3.12) with k replaced by $k-1$.*

Proof. We set $w = R_{u_{k-1}} v$. Clearly w possesses the correct initial values since $w(0) = u_{k-1}$ and $w'(0) = v(0) = u_k$. By a straightforward (albeit lengthy) computation we show that

$$\begin{aligned} (\mathcal{C}w')' = f^{(k-1)} &- ((k-1)\mathcal{C}' + \mathcal{B})w' \\ &- \sum_{j=0}^{k-1} \left[\binom{k-1}{j} (\mathcal{A}^{(j)} + \mathcal{Q}^{(j)}) + \binom{k-1}{j+1} \mathcal{B}^{(j+1)} \right. \\ &\left. + \binom{k}{j+2} \mathcal{C}^{(j+2)} \right] \bigcirc_{l=k-1-j}^{k-2} R_{u_l} w \end{aligned} \tag{3.14}$$

holds in the $L^2(I;V^*)$-sense. We start by employing the fundamental theorem of calculus to the sum in equation (3.14), which we denote by Γ. By turning to (3.11), we discover that $\Gamma(0)$ can be rewritten using $C(0)u_{k+1}$, since

$$\Gamma(0) = -C(0)u_{k+1} + f^{(k-1)}(0) - (kC'(0) + B(0))u_k.$$

Using our product rule (Lemma 2.3) we can reshape Γ' to obtain

$$\begin{aligned}\Gamma' = \sum_{j=0}^{k-1} \Bigg[&\binom{k-1}{j}(\mathcal{A}^{(j+1)} + \mathcal{Q}^{(j+1)}) + \binom{k-1}{j+1}\mathcal{B}^{(j+2)} \\ &+ \binom{k}{j+2}\mathcal{C}^{(j+3)}\Bigg]\left(\bigcirc_{l=k-1-j}^{k-1} R_{u_l} v\right) \\ + \sum_{j=0}^{k-1} \Bigg[&\binom{k-1}{j}(\mathcal{A}^{(j)} + \mathcal{Q}^{(j)}) + \binom{k-1}{j+1}\mathcal{B}^{(j+1)} \\ &+ \binom{k}{j+2}\mathcal{C}^{(j+2)}\Bigg]\left(\bigcirc_{l=k-j}^{k-1} R_{u_l} v\right).\end{aligned}$$

Shifting the index of the first sum and combining the result with the second sum (via summation rules for binomial coefficients) shows

$$\begin{aligned}\Gamma' = -(k\mathcal{C}'' + \mathcal{B}')(v) + \sum_{j=0}^{k} \Bigg[&\binom{k}{j}(\mathcal{A}^{(j)} + \mathcal{Q}^{(j)}) + \binom{k}{j+1}\mathcal{B}^{(j+1)} \\ &+ \binom{k+1}{j+2}\mathcal{C}^{(j+2)}\Bigg]\left(\bigcirc_{l=k-j}^{k-1} R_{u_l} v\right).\end{aligned}$$

This is already pretty close to the left-hand side of (3.12a), which we abbreviate by $L^{[k]}v$. Replacing the sum in (3.14) by $\Gamma(0) + \int_0^t \Gamma'(s)\,ds$ and also applying the fundamental theorem to the remaining parts allows us to conclude the proof with

$$\begin{aligned}&f^{(k-1)}(t) - ((k-1)C'(t) + B(t))w'(t) - \Gamma(t) \\ &= C(0)u_{k+1} + C'(0)u_k + \int_0^t f^{(k)}(s) + (\mathcal{C}'v)'(s) - \left(L^{[k]}v - (\mathcal{C}'v)'\right)(s)\,ds \\ &= C(0)u_{k+1} + C'(0)u_k + \int_0^t (\mathcal{C}v')'(s) + (\mathcal{C}'v)'(s)\,ds \\ &= C(0)u_{k+1} + C'(0)u_k + \int_0^t (\mathcal{C}v)''(s)\,ds = (\mathcal{C}v)'(t) = (\mathcal{C}w')'(t). \qquad \square\end{aligned}$$

Finally, we can state the regularity theorem and prove it using a compact induction argument.

Theorem 3.10. *Let* $k \in \mathbb{N}$ *and* $a_0, c_0 > 0$. *We assume that the operators satisfy* $A \in W^{k+1,\infty}(I;\mathcal{L}^{sa}(V,V^*))$, $B \in W^{k,\infty}(I;\mathcal{L}(H))$, $C \in W^{k+1,\infty}(I;\mathcal{L}^{sa}(H))$ *and* $Q \in W^{k,\infty}(I;\mathcal{L}(V,H))$ *with* $A(t) \in \mathcal{L}^{sa}_{a_0}(V,V^*)$ *and* $C(t) \in \mathcal{L}^{sa}_{c_0}(H)$ *for almost all* $t \in I$. *Further, let* $f \in H^k(I;H) \cup H^{k+1}(I;V^*)$, $u_j \in V$ *for*

$j = 0, \dots, k$ *and* $u_{k+1} \in H$. *Then the unique solution* u *of problem* (3.1) *lies in* $H^k(I;V) \cap H^{k+1}(I;H)$ *and satisfies the energy estimate*

$$\|u\|^2_{W^{k,\infty}(I;V)} + \|u^{(k+1)}\|^2_{L^\infty(I;H)} \leqslant \Lambda \left(\sum_{j=0}^{k} \|u_j\|^2_V + \|u_{k+1}\|^2_H + \|f\|^2 \right), \tag{3.15}$$

where f *is measured in either the* $H^k(I;H)$*- or the* $H^{k+1}(I;V^*)$*-norm and* $\Lambda > 0$ *is a constant depending continuously on* c_0^{-1}, a_0^{-1} *as well as the operators* A, B, C *and* Q.

Proof. Induction over $k \in \mathbb{N} \cup \{0\}$, where the hypothesis for $k \geqslant 1$ is that the assertion as given in the theorem holds and that $u^{(k)}$ is the unique solution of the auxiliary problem (3.12).[1] This is also the hypothesis for $k = 0$, but with the additional assumptions $Q \in W^{1,\infty}(I;\mathcal{L}(V,H)) \cup L^\infty(I;\mathcal{L}(H))$ and $B \in W^{1,\infty}(I;\mathcal{L}(H))$.

The case $k = 0$ is covered by Theorem 2.10. We assume that the hypothesis holds for $k-1$ and that the requirements for k are met. Through Corollary 3.8 we know that a solution $v \in L^2(I;V) \cap H^1(I;H)$ to the auxiliary problem (3.12) exists. Due to Lemma 3.9 we are able to conclude that $R_{u_{k-1}}v$ satisfies (3.12) for $k-1$, which (by assumption) is uniquely solved by $u^{(k-1)}$. Hence, $v = (R_{u_{k-1}}v)' = u^{(k)}$ and in particular, v is unique. By adding the estimate (3.13) that is satisfied by $u^{(k)}$ to inequality (3.15) with k replaced by $k-1$, we obtain validity of the latter for k. □

Had we iterated the "original" approach from Theorem 3.5 (by moving $\mathcal{A}^{(j)}u^{(k-j)}$ to the right-hand side), then we would have obtained a result similar to Theorem 3.10, where $A \in W^{k,\infty}(I;\mathcal{L}(V,V^*))$ with the restriction $A' \in W^{k-1,\infty}(I;\mathcal{L}(V,H))$. Thus, the price we had to pay to allow $A'(t) \in \mathcal{L}(V,V^*)$ was another order of differentiability in time for A.

3.3 CONCLUSIONS

In the preceding sections we have derived two regularity results for the evolution problem (3.1): Theorem 3.1, obtained using Lions' techniques, and Theorem 3.10, obtained by differentiation in time. Both theorems make the same assumptions on the operators and how smooth they have to be with respect to the time variable. A subtle difference can be found in the energy estimates that they contain; Lions obtains L^2-estimates for u and its derivatives, whereas the other technique yields slightly better L^∞-type estimates. The only *significant* difference is the fact that Lions' result requires homogeneous initial values and a vanishing right-hand side at the initial time in order to avoid compatibility conditions. They appear in Theorem 3.10 through the requirement that u_j (as defined

1 We include $k = 0$ because we utilize the uniqueness of $u = u^{(0)}$ in the proof. Otherwise the case $k = 1$ would have to employ the induction argument as well.

in (3.11)) must belong to V for $j = 0, \dots, k$ and that u_{k+1} has to be an element of H. These conditions take the form

$$u_0,\, u_1 \in V,\; C(0)u_2 = f(0) - [C'(0) + B(0)]u_1 - [A(0) + Q(0)]u_0 \in H$$

in the case $k = 1$ and

$$\begin{aligned} &u_0,\, u_1,\, u_2 = C(0)^{-1}\big(f(0) - [C'(0) + B(0)]u_1 - [A(0) + Q(0)]u_0\big) \in V,\\ &C(0)u_3 = f'(0) - [2C'(0) + B(0)]u_2 - [A(0) + Q(0) + B'(0) + C''(0)]u_1\\ &\qquad\qquad - [A'(0) + Q'(0)]u_0 \quad \in H \end{aligned}$$

when $k = 2$. When we choose the operators as time-independent, then these compatibility conditions are identical to those that prominently appear in the literature, for example in [Wlo87]. They encode "spatial" regularity of the data, the operators and their time derivatives at the initial time. It is hard to see this in the abstract framework, because we have no detailed knowledge about the Hilbert spaces V and H. However, this is not impossible, for example it is feasible to talk about the domain

$$D(A(t)) := \{\, \varphi \in V \mid A(t)\varphi \in H \,\}$$

of $A(t)$. Obviously one can also elaborate on this and for instance look at $\varphi \in V$ with $A(t)\varphi \in V$ or $A(t)\varphi \in \mathcal{D}(A(t))$. The space $D(A(t))$ indirectly appears in the regularity results: If we are in the position to apply Theorem 3.10 for $k = 1$, then $(\mathcal{C}u')' \in L^2(I; H)$. If we further suppose that $f(t) \in H^1(I; H)$, then from the evolution equation it follows that $\mathcal{A}u \in L^2(I; H)$, i.e. $u(t) \in D(A(t))$ for almost all $t \in I$. Moreover, the compatibility condition $C(0)u_2 \in H$ is satisfied if $u_0 \in D(A(0))$.

In applications, the space $D(A(t))$ involves well-defined second derivatives. We give an example that underlines this fact and also shows how the compatibility conditions can be fulfilled in practice. For this we continue examples 2.4 and 2.11.

Example 3.11. Let $\Omega \subset \mathbb{R}^d$ be a domain. We know that the weak formulation of

$$\begin{gathered} u''(t,x) - \operatorname{div}(a(t,x)\nabla u(t,x)) = f(t,x) \quad t \in (0,T),\ x \in \Omega\\ u = 0 \text{ on } (0,T) \times \partial\Omega,\ u(0,\cdot) = u_0,\ u'(0,\cdot) = u_1 \end{gathered}$$

is equivalent to the evolution problem $u'' + \mathcal{A}u = f$, $u(0) = u_0$ and $u'(0) = u_1$, with $\langle A(t)\varphi, \psi\rangle = \int_\Omega a(t)\nabla\varphi \cdot \nabla\psi \,dx$, $V = H^1_0(\Omega)$ and $H = L^2(\Omega)$. A function $v \in H^1_0(\Omega)$ lies in the domain $D(A(t))$ of $A(t)$ if and only if there exists $w \in L^2(\Omega)$ such that $\langle A(t)v, \psi\rangle = (w, \psi)$ holds for all $\psi \in H^1_0(\Omega)$. By definition this means that $a(t)\nabla v$ has a well-defined divergence that is equal to $w \in L^2(\Omega)$. Notably, if $a(t) \in W^{1,\infty}(\Omega)$ then we discover $H^2(\Omega) \cap H^1_0(\Omega) \subset D(A(t))$.

Now we turn to the compatibility conditions. According to (3.11), the initial value u_j is for $j \geqslant 2$ equal to

$$u_j = f^{(j-2)}(0) - \sum_{l=0}^{j-2} \binom{j-2}{l} A^{(l)}(0)u_{j-2-l}.$$

To make the assumptions $u_j \in H_0^1(\Omega)$ $(j = 0, \dots, k)$, $u_{k+1} \in L^2(\Omega)$ valid, not every summand has to belong to $H_0^1(\Omega)$ or $L^2(\Omega)$; it is acceptable if the $A^{(l)}(0)u_{j-2-l}$ and $f^{(j-2)}(0)$ "fit together" so that their sum is regular enough (hence the name *compatibility* conditions). However, to give a sufficient criterion that is not recursively defined and makes the conditions valid, we will simply require this smoothness for each summand. Let $H_{00}^m(\Omega)$ denote $H^m(\Omega) \cap H_0^1(\Omega)$ for $m \geqslant 1$ and $H_{00}^0(\Omega) := L^2(\Omega)$. A short calculation shows that $u_j \in H^m(\Omega)$ $(m \geqslant 0, j \geqslant 2)$ if for $l = 0, \dots, j-2$

$$f^{(j-2)}(0) \in H^m(\Omega), \quad u_l \in H^{m+2}(\Omega) \quad \text{and} \quad a^{(l)}(0) \in W^{m+1,\infty}(\Omega)$$

holds. Thus, to ensure that the compatibility conditions for $k \in \mathbb{N}$ are satisfied, it is sufficient to require $u_0 \in H_{00}^{k+1}(\Omega)$, $u_1 \in H_{00}^k(\Omega)$,

$$a^{(l)}(0) \in W^{k-l,\infty}(\Omega) \quad \text{and} \quad f^{(l)}(0) \in H_{00}^{k-l-1}(\Omega)$$

for $l = 0, \dots, k-1$. In particular, $a(0)$ must be k-times weakly differentiable with respect to the spatial coordinates.

The regularity theorems do not only enable us to analyze the inverse problems that will be introduced in the next chapter, but have another very useful implication: Originally, we were not able to give theory for equations like $C(t)u''(t) + A(t)u(t) = f(t)$, since even the problem statement for this case is difficult (both $u'(0)$ and $(\mathcal{C}u')(0)$ are not well-defined). However, for $u \in H^2(I; H)$ this equation is equivalent to $(\mathcal{C}u')' - \mathcal{C}'u' + \mathcal{A}u = f$. We can solve the latter equation, and are now able to show that u is sufficiently regular to warrant this application of the product rule.

Corollary 3.12. *Let $k \in \mathbb{N}$ and $a_0, c_0 > 0$. We assume that operators $A \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*))$, $B \in W^{k,\infty}(I; \mathcal{L}(H))$, $C, D \in W^{k+1,\infty}(I; \mathcal{L}(H))$ and $Q \in W^{k,\infty}(I; \mathcal{L}(V, H))$ are given, with $A(t) \in \mathcal{L}_{a_0}^{sa}(V, V^*)$ and $D(t)C(t) \in \mathcal{L}_{c_0}^{sa}(H)$ for almost all $t \in I$. Further, let $f \in H^k(I; H) \cup H^{k+1}(I; V^*)$ and suppose that $u_j \in V$ $(j = 0, \dots, k)$ and $u_{k+1} \in V$, where the u_j are as in (3.11), but with C replaced by DC and B replaced by $B - D'C$. Then there exists a unique solution $u \in H^k(I; V) \cap H^{k+1}(I; H)$ of the equation*

$$\mathcal{D}(\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{A} + \mathcal{Q})u = f$$

that also possesses the initial values $u(0) = u_0$ and $(\mathcal{D}\mathcal{C}u')(0) = u_1$. Moreover, this solution satisfies the estimate (3.15).

Proof. With the assumptions made in the assertion, we can apply Theorem 3.10 to see that there is a unique $u \in H^k(I; V) \cap H^{k+1}(I; H)$ that solves

$$(\mathcal{D}\mathcal{C}u')' + (\mathcal{B} - \mathcal{D}'\mathcal{C})u' + (\mathcal{A} + \mathcal{Q})u = f$$

and further satisfies the initial conditions and estimate (3.15). We have $u \in H^2(I; H)$, hence Lemma 2.3 yields $(\mathcal{D}\mathcal{C}u')' = \mathcal{D}(\mathcal{C}u')' + \mathcal{D}'\mathcal{C}u'$. □

4 ABSTRACT OPERATOR FRAMEWORK

The preceding chapters have laid the foundation for the definition of forward operators to inverse problems that are related to the evolution problem (2.6), that is

$$(\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{A}+\mathcal{Q})u = f \quad \text{in } L^2(I;V^*), \tag{4.1a}$$

$$u(0) = u_0 \;\text{ and }\; (\mathcal{C}u')(0) = u_1. \tag{4.1b}$$

In Example 2.4 we have already seen how partial differential equations can be restated in this formulation, and that coefficients in these equations will influence one or more of the operators A, B, C and Q. These operators then determine the solution u of the PDE. In view of applications, it is more realistic to subsequently apply a measurement operator to u. Depending on the experimental setup, this operator could for example restrict u to the boundary or to sensor locations. These consideration yield a natural decomposition of the problem's forward operator F into a *value* operator P, that maps the unknowns onto (A, B, C, Q), followed by the solution operator S of (4.1) and a measurement operator Ψ. This is sketched in Figure 4.1.

This approach has the striking advantage that only two of the three operators are problem-dependent, and S, the most difficult one to analyze, is not one of them. If we assume the measurement operator Ψ and the operator P that maps the unknown parameter (or parameters) to the tuple (A, B, C, Q) to be Fréchet-differentiable, then many properties of the forward operator F can be derived by analyzing these operators separately. These properties include Fréchet-differentiability of F and the shape of the corresponding adjoints. They are therefore crucial for the numerical implementation of Newton-based regularization methods.

A related setup to Figure 4.1, which we will use in some numerical experiments in Chapter 8, is that not only a single u is measured, but that the same unknowns are used to generate multiple wave fields $u_1, \dots, u_m$ with a different right-hand side f in the equation. This results in a vector-valued solution operator $\mathbf{S} = (S_1, \dots, S_m)$. As long as the count m of measured fields is finite, the corresponding problems can be dealt with in the same fashion as S.

$$\begin{tikzcd}
\text{unknowns} \arrow[r, "P"] \arrow[rrr, bend right, "F := \Psi \circ S \circ P"'] & (A, B, C, Q) \arrow[r, "S"] & u \arrow[r, "\Psi"] & \text{measurements}
\end{tikzcd}$$

Figure 4.1: Decomposition of a possible forward operator F

In this chapter, we will discuss the problem-independent operator S. Our main idea is to consider S itself as the forward operator of an inverse problem, in which the operators in (4.1a) represent the unknowns. Since this operator is not directly related to a particular partial differential equation, we designate this as an "abstract" inverse problem. This sentiment is reinforced by the fact that even after fixing spaces V and H, it is most certainly neither relevant for the application nor numerically feasible to reconstruct the operators A, B, C, Q without making further restrictions on their structure first.

We begin in the first section with a precise definition of S. Since we aim to show its differentiability in Section 4.2, it is crucial to ensure that the domain of definition of S is an open subset of a Banach space. We close the abstract theory by showing local ill-posedness of the abstract inverse problem, and also the ill-posedness of its linearizations in Section 4.3.

We remark that most of the results of this chapter have also been presented in [Ger19].

4.1 WELL-POSEDNESS OF THE FORWARD OPERATOR

As in Chapter 2, problem (4.1) is stated in the Gelfand triple $V \subset H \subset V^*$, consisting of separable Hilbert spaces V and H that provide a compact and dense embedding $V \hookrightarrow H$. Without loss of generality we still assume $\|\cdot\|_H \leqslant \|\cdot\|_V$, and that the time interval I is given as $[0, T]$ for some fixed $T > 0$. We will also continue to use the notation that was introduced on page 6.

Let $c_0, a_0 > 0$, $f \in L^2(I; H) \cup H^1(I; V^*)$, $u_0 \in V$ and $u_1 \in H$ be fixed throughout the chapter. According to Theorem 2.10, the operator S that maps the tuple (A, B, C, Q) of unknowns to the solution u of (4.1) is well-defined as $S\colon \widetilde{D}(S) \subset X \to Y$, where the Banach spaces X and Y are defined as

$$
\begin{aligned}
X &:= W^{1,\infty}(I; \mathcal{L}^{sa}(V, V^*)) \times W^{1,\infty}(I; \mathcal{L}(H)) \\
&\qquad \times W^{1,\infty}(I; \mathcal{L}^{sa}(H)) \times W^{1,\infty}(I; \mathcal{L}(V, H)), \\
Y &:= L^\infty(I; V) \cap W^{1,\infty}(I; H)
\end{aligned}
$$

and the domain $\widetilde{D}(S)$ of S is set to

$$
\widetilde{D}(S) = \Big\{ (A, B, C, Q) \in X \,\Big|\, A(t) \in \mathcal{L}^{sa}_{a_0}(V, V^*) \text{ and } C(t) \in \mathcal{L}^{sa}_{c_0}(H) \\
\text{for almost all } t \in I \Big\}.
$$

Note that in the case $Q(t) \in \mathcal{L}(H)$, we could omit the differentiability assumption on Q in the definition of X without impacting any of the results to come. For the definition of the Banach space $\mathcal{L}^{sa}(Z, Z^*)$ of self-adjoint, bounded linear operators on a normed space Z, and its subset $\mathcal{L}^{sa}_\alpha(Z, Z^*)$ of operators that additionally are α-coercive, we refer to Definition 2.7 on page 14.

As was already hinted at in the introduction of this chapter, for a proper analysis of the differentiability of S we need it to be defined on an open subset of a Banach space. Unfortunately, $\widetilde{D}(S)$ is not open in X because $\mathcal{L}^{sa}_{\alpha}(Z, Z^*)$ is closed. The interior of this set is not obtained by simply using "$>$" instead of "$\geqslant$" in its definition (and restrict the assumption to $Z \setminus \{0\}$) because in general this set is also not open. We will prove this fact by giving a simple counter-example.

Example 4.1. Let $\alpha \in [0, 1)$ and $Z := L^2([\alpha, 1])$. We choose to identify Z^* with Z and define $F \in \mathcal{L}^{sa}(Z)$ for $v \in Z$ through

$$F(v) = (x \mapsto xv(x)) \in Z.$$

This means $(F(v), v) = \int_{\alpha}^{1} xv(x)^2 \, dx > \alpha \|v\|_Z^2$ for all $v \neq 0$, in particular $F \in \mathcal{L}^{sa}_{\alpha}(Z)$. We further define the family of operators

$$F_{\varepsilon}(v) := (x \mapsto (x - \varepsilon)v(x))$$

with $\varepsilon > 0$. By $\mathbb{1}_{\Omega}$ we denote the characteristic function of Ω, see that

$$\left(F_{\varepsilon}(\mathbb{1}_{[\alpha,\alpha+\varepsilon]}), \mathbb{1}_{[\alpha,\alpha+\varepsilon]}\right) = \int_{\alpha}^{\alpha+\varepsilon} x - \varepsilon \, dx = (\alpha - \varepsilon/2)\varepsilon < \varepsilon = \left\|\mathbb{1}_{[\alpha,\alpha+\varepsilon]}\right\|_Z^2$$

and conclude $F_{\varepsilon} \notin \mathcal{L}^{sa}_{\alpha}(Z)$. Furthermore, $\|F - F_{\varepsilon}\|_{\mathcal{L}(Z)} \leqslant \varepsilon$, so $F_{\varepsilon} \to F$ when $\varepsilon \to 0$. Since F can be approximated by a sequence in $\mathcal{L}^{sa}(Z) \setminus \mathcal{L}^{sa}_{\alpha}(Z)$, it cannot belong to the interior of $\mathcal{L}^{sa}_{\alpha}(Z)$.

Fortunately for us, the interior of $\mathcal{L}^{sa}_{\alpha}(Z, Z^*)$ is not empty, and we can give a short formula for it.

Lemma 4.2. *Let Z be a Hilbert space and $\alpha > 0$. The interior of $\mathcal{L}^{sa}_{\alpha}(Z, Z^*)$ is given by*

$$\mathcal{L}^{sa}_{\alpha}(Z, Z^*)^{\circ} = \bigcup_{\varepsilon > 0} \mathcal{L}^{sa}_{\alpha+\varepsilon}(Z, Z^*).$$

Proof. We set $M = \bigcup_{\varepsilon>0} \mathcal{L}^{sa}_{\alpha+\varepsilon}(Z, Z^*)$ and show that it is the largest open subset of $\mathcal{L}^{sa}_{\alpha}(Z, Z^*)$. It is obvious that $M \subset \mathcal{L}^{sa}_{\alpha}(Z, Z^*)$. We continue by proving that M is open. Let $G \in M$, which means there is $\varepsilon_0 > 0$ such that $G \in \mathcal{L}^{sa}_{\alpha+\varepsilon_0}(Z, Z^*)$. For every $F \in B(G, \varepsilon_0/2)$ (open ball around G with respect to the operator norm) and $v \in Z$ we have

$$\begin{aligned}\langle Fv, v\rangle &= \langle Gv, v\rangle + \langle (F - G)v, v\rangle \geqslant (\alpha + \varepsilon_0)\|v\|^2 - \|F - G\|\|v\|^2 \\ &\geqslant (\alpha + \varepsilon_0/2)\|v\|^2,\end{aligned}$$

which means that $F \in \mathcal{L}^{sa}_{\alpha+\varepsilon_0/2}(Z, Z^*) \subset M$, hence M must be open. As a last step we show that every $G \in \mathcal{L}^{sa}_{\alpha}(Z, Z^*) \setminus M$ can be approximated by operators that belong to $\mathcal{L}^{sa}(Z, Z^*) \setminus \mathcal{L}^{sa}_{\alpha}(Z, Z^*)$. Since $G \notin M$ there exists a sequence $(v_k)_{k\in\mathbb{N}} \subset Z$ with $\langle Gv_k, v_k\rangle = (\alpha + 1/(2k))\|v_k\|^2$. We set $G_k = G - 1/k \, I_{Z\to Z^*}$, where $I_{Z\to Z^*}$ denotes the canonical embedding of the Hilbert space Z in its dual space. It is easy to verify that $G_k \to G$ for $k \to \infty$, as well as $\langle G_k v_k, v_k\rangle = (\alpha - 1/(2k))\|v_k\|^2$, so none of the G_k belong to $\mathcal{L}^{sa}_{\alpha}(Z, Z^*)$. $\square$

In view of the previous lemma, we replace $\widetilde{D}(S)$ by its interior

$$D(S) := \Big\{ (A,B,C,Q) \in X \;\Big|\; A(t) \in \mathcal{L}^{sa}_{a_0+\varepsilon}(V,V^*) \text{ and } C(t) \in \mathcal{L}^{sa}_{c_0+\varepsilon}(H) \\ \text{for almost all } t \in I \text{ for some } \varepsilon > 0 \Big\}.$$

Since X is equipped with L^∞-type norms in the time variable, it is easy to check that $\widetilde{D}(S)^\circ = D(S)$. Thus, we obtain the solution operator

$$S\colon D(S) \subset X \to Y, \quad (A,B,C,Q) \mapsto u$$

to problem (4.1) that is defined on an open subset of the Banach space X.

Our regularity results from Chapter 3 also permit the definition of a variant of S which maps into a subspace $Y^{(k)}$ $(k \in \mathbb{N})$ of Y consisting of more regular wave fields u, i.e.

$$Y^{(k)} := W^{k,\infty}(I;V) \cap W^{k+1,\infty}(I;H). \tag{4.2}$$

From Theorem 3.10 we see that if we wish to obtain $S(A,B,C,Q) \in Y^{(k)}$, the operators must belong to the subspace

$$X^{(k)} := W^{k+1,\infty}(I;\mathcal{L}^{sa}(V,V^*)) \times W^{k,\infty}(I;\mathcal{L}(H)) \\ \times W^{k+1,\infty}(I;\mathcal{L}^{sa}(H)) \times W^{k,\infty}(I;\mathcal{L}(V,H)) \tag{4.3}$$

of X. Naturally, the tuple (A,B,C,Q) has to be an element of $D(S)$ in order to provide coercivity of A and C. Moreover, we have to assume $f \in H^k(I;H)$ or $f \in H^{k+1}(I;V^*)$ and that the compatibility conditions

$$u_0,\dots,u_k \in V, \quad u_{k+1} \in H$$

are satisfied. The u_l are given for $l \geqslant 2$ by equation (3.11), which we repeat here for the reader's convenience; it reads

$$C(0)u_{l+2} = f^{(l)}(0) - ((l+1)C'(0) + B(0))u_{l+1} \\ - \sum_{j=0}^{l} \Bigg[\binom{l}{j}(A^{(j)}(0) + Q^{(j)}(0)) \\ + \binom{l}{j+1} B^{(j+1)}(0) + \binom{l+1}{j+2} C^{(j+2)}(0) \Bigg] u_{l-j}.$$

The compatibility conditions have been discussed in Section 3.3 in more detail. There we have discovered that they tend to become unwieldy in practice, even for trivial $C = \mathrm{Id}$ and $B = Q = 0$. In general, the set of all operators A, B, C and Q that satisfy the compatibility conditions is not a linear space. Therefore we cannot put them into the definition of X. Furthermore, it is also not constructive to integrate them into $D(S)$, because then it will cease to be an open subset of X. Thus, they have to be linearized somehow. We do not want to go into the detail how this can be achieved in the general case $k \in \mathbb{N}$, and instead only motivate this by demonstrating it for $k = 1$. In this case we have to ensure $u_0, u_1 \in V$ and $C(0)u_2 \in H$. The latter is satisfied if

$$f(0) - (C'(0) + B(0))u_1 - (A(0) + Q(0))u_0 \in H,$$

which in turn is valid if $f(0) \in H$ and $u_0 \in D(A(0))$, which can be incorporated into $X^{(1)}$. Even for $k = 2$ this method (assuming that every summand is regular enough on its own) becomes significantly more complicated because u_j with $j \geqslant 2$ depends nonlinearly on A. Thus, the set of all A that fulfill a condition like $u_2 \in D(A(0))$ is not a linear space. In consequence, this condition has to be deconstructed further by inserting the definition of u_2 and analyzing each arising term individually.

The subsequent analysis is not impacted by the way that the compatibility conditions are treated, as long as $D(S)$ remains an open set in a (possibly different) Banach space $X^{(k)}$. Moreover, we do not profit from having found the most "intricate" way of satisfying them that yields the minimal additional assumptions on the data and the operators. Therefore we choose the simplest option available by requiring homogeneous initial values $u_0 = u_1 = 0$, $f^{(j)}(0) = 0$ for $j = 0, \dots, k-2$ (provided that $k \geqslant 2$) and $f^{(k-1)}(0) \in H$ (provided that $k \geqslant 1$). In this way, $u_j = 0 \in V$ for $j = 0, \dots, k$ and $u_{k+1} = C(0)^{-1} f^{(k-1)}(0) \in H$. We thereby entirely avoid additional constraints that have to appear in the space X. Under these assumptions on f and the initial values, we can view S for all $k \in \mathbb{N}$ as the operator

$$S \colon D(S) \cap X^{(k)} \subset X^{(k)} \to Y^{(k)}$$

with Banach spaces $X^{(k)}$ and $Y^{(k)}$ as defined in (4.3) and (4.2). We note that $D(S) \cap X^{(k)}$ is an open subset of $X^{(k)}$. We extend this definition to $k = 0$ by setting[1] $X^{(0)} := X$ and $Y^{(0)} := Y$.

To avoid having to repeat the conditions that f has to fulfill in every assertion, we define the set

$$\mathcal{F}^{(k)} := \Big\{ f \in H^k(I;H) \cup H^{k+1}(I;V^*) \;\Big|\; f^{(k-1)}(0) \in H \text{ if } k \geqslant 1 \\ \text{and } f^{(j)}(0) = 0 \text{ for all } j = 0, \dots, k-2 \text{ if } k \geqslant 2 \Big\} \tag{4.4}$$

of admissible right-hand sides.

We would like to close the section with the remark that the conditions on f resemble the assumptions of Lions' regularity result (Theorem 3.1 on page 25). The difference is that we only require $f^{(k-1)}(0) \in H$ instead of $f^{(k-1)}(0) = 0$. If one were to assume the latter, then the more regular variant of S could also have been set up by employing Theorem 3.1 instead of Theorem 3.10, albeit with the slightly larger image space $Y^{(k)} := H^k(I;V) \cap H^{k+1}(I;H)$.

1 Note that X is not the same space that one would obtain by setting $k = 0$ in (4.3), because X also contains assumptions on Q and B that guarantee the uniqueness of the solution u.

4.2 FRÉCHET-DIFFERENTIABILITY

One of the main approaches for the numerical solution of nonlinear inverse problems is by means of the linearization of their forward operators. How these derivatives are used depends on the regularization method. For example, it is of direct interest in Newton-methods (e.g. REGINN), where the nonlinear problem is reduced to a sequence of linear subproblems. In contrast, for Tikhonov regularization it might be used as a means to compute a descent direction for a nonlinear functional. The derivative of a forward operator that is based on S (like depicted in Figure 4.1) is easily calculated if the derivative of S is known.

We use the term "differentiability" as a synonym to Fréchet-differentiability. Weaker notions of derivatives (e.g. in the sense of Gateaux) would not reduce the assumptions on the unknown operators, at least in the approach presented here. This is due to the fact that we make heavy use of the energy estimates we derived in the previous chapter (Theorems 2.10 and 3.10), thus our results will only depend on the norm of the linearization argument, not its direction. Indeed, for time-independent parameters and in the context of the acoustic wave equation, the Gateaux-derivative turns out to also be the Frechet-derivative (see [KR14a; Sto01]).

We will begin with the partial derivatives of S, because they show which of the four unknowns cause difficulties. At the end of the section we can use them to construct the total derivative of S. As a starting point, we need a hypothesis what the derivative of S with respect to the operators could be. We will use the operator $B \in W^{1,\infty}(I;\mathcal{L}(H))$ to demonstrate the approach. Let $p = (A, B, C, Q) \in D(S)$ and $u = S(p)$ be the solution of problem (4.1). Given the linearization argument $h \in W^{1,\infty}(I;\mathcal{L}(H))$, we seek a characterization of $u_h := \partial_B S(p)[h]$. We formally linearize the evolution equation $(\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{A} + \mathcal{Q})u = f$ with respect to B. Since the evolution equation is already linear, only the term $\mathcal{B}u'$ changes; in the remaining terms of the left-hand side we can simply replace u with u_h. Hence, the result is

$$(\mathcal{C}u_h')'(t) + B(t)u_h'(t) + h(t)u'(t) + (A(t) + Q(t))u_h(t) = 0, \quad t \in I,$$

which means that u_h solves the same evolution equation as u, but with the right-hand side $t \mapsto -h(t)u'(t)$, i.e. the linearization argument *applied to* the (negative) derivative of the solution of the forward problem. We make the observation that the initial values (4.1b) do not depend on B and therefore conclude that u_h and $\mathcal{C}u_h'$ should vanish at the initial time. The same argument can be made for the operators A and Q, and under the additional assumption $u_1 = 0$ also for C. We will see that we have to require the latter anyway. Therefore we make the hypothesis that for each symbol $x \in \{A, B, C, Q\}$, the derivative $u_h = \partial_x S(p)[h]$ solves the equation

$$(\mathcal{C}u_h')' + \mathcal{B}u_h' + (\mathcal{A} + \mathcal{Q})u_h = g_x(u)[h] \tag{4.5a}$$

in $L^2(I;V^*)$ and possesses homogeneous initial values

$$u_h(0) = 0, \quad (\mathcal{C}u_h')(0) = 0. \tag{4.5b}$$

Naturally, both the function space for h and the specific shape of the right-hand side depend on the symbol x. The auxiliary functions

$$\begin{aligned} g_A(v)[H] &= -\mathcal{H}[v] = -H(\cdot)[v(\cdot)], & g_C(v)[H] &= -(\mathcal{H}[v'])', \\ g_B(v)[H] &= -\mathcal{H}[v'], & g_Q(v)[H] &= -\mathcal{H}[v] \end{aligned} \tag{4.6}$$

help us to achieve a closed presentation. Note that the right-hand sides for A and Q are the same, but we will have g_A and g_Q map between different spaces.

With (4.5) we have obtained an evolution problem for u_h, which we might be able to solve using the well-posedness Theorem 2.10. To obtain a unique $u_h \in Y^{(0)}$, we have to ensure that g_x maps either into $L^2(I;H)$ or $H^1(I;V^*)$. The natural choice of domains and ranges for the g_x that accomplish this are

$$\begin{aligned} g_A(\cdot)[\cdot]&\colon H^1(I;V) \times W^{1,\infty}(I;\mathcal{L}(V,V^*)) \to H^1(I;V^*), \\ g_B(\cdot)[\cdot]&\colon H^1(I;H) \times L^{\infty}(I;\mathcal{L}(H)) \to L^2(I;H), \\ g_C(\cdot)[\cdot]&\colon H^2(I;H) \times W^{1,\infty}(I;\mathcal{L}(H)) \to L^2(I;H), \\ g_Q(\cdot)[\cdot]&\colon L^2(I;V) \times L^{\infty}(I;\mathcal{L}(V,H)) \to L^2(I;H). \end{aligned}$$

This way we obtain continuous bilinear forms, e.g.

$$g_A(\cdot)[\cdot] \in \mathcal{L}(H^1(I;V),\ \mathcal{L}(W^{1,\infty}(I;\mathcal{L}(V,V^*)),\ H^1(I;V^*))),$$

and we can already deduce that $u \in Y^{(0)}$ is not enough to apply g_A or g_C to it. In these cases we need at least $u \in Y^{(1)}$ to make u_h well-defined. If we also want to ensure higher regularity of u_h, we have to use the continuous bilinear forms

$$\begin{aligned} g_A(\cdot)[\cdot]&\colon H^{k+1}(I;V) \times W^{k+1,\infty}(I;\mathcal{L}(V,V^*)) \to H^{k+1}(I;V^*), \\ g_B(\cdot)[\cdot]&\colon H^{k+1}(I;H) \times W^{k,\infty}(I;\mathcal{L}(H)) \to H^k(I;H), \\ g_C(\cdot)[\cdot]&\colon H^{k+2}(I;H) \times W^{k+1,\infty}(I;\mathcal{L}(H)) \to H^k(I;H), \\ g_Q(\cdot)[\cdot]&\colon H^k(I;V) \times W^{k,\infty}(I;\mathcal{L}(V,H)) \to H^k(I;H), \end{aligned}$$

resulting in $u_h \in Y^{(k)}$ ($k \in \mathbb{N}$), as long as $p = (A,B,C,Q)$ belongs to $D(S) \cap X^{(k)}$ and u is regular enough. We see that in this case we can insert $u \in Y^{(k+1)}$ into g_A and g_C, whereas $u \in Y^{(k)}$ is already sufficient for g_B and g_Q.

The discussion above leads to the impression that $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$ might be differentiable in B and Q, and that $S\colon D(S) \cap X^{(k)} \to Y^{(k-1)}$ is additionally differentiable in A and C. However, this is not the case. Up to now we have only established existence and uniqueness of u_h. For the proof that u_h indeed describes the Fréchet-derivative of S we need another ingredient, namely that S is locally Lipschitz continuous. For A and C this will cost another order of regularity, as the following Lemma demonstrates.

Lemma 4.3. *Let $k \in \mathbb{N} \cup \{0\}$ and $f \in \mathcal{F}^{(k)}$. If $k \neq 0$ then we also assume $u_0 = u_1 = 0$. Under these conditions,* *(i) the map* $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$ *is locally Lipschitz continuous with respect to* B *and* Q*, and*

(ii) provided that $k \geqslant 1$*, the map* $S\colon D(S) \cap X^{(k)} \to Y^{(k-1)}$ *is locally Lipschitz continuous.*

In both cases the Lipschitz constants continuously depend on the argument measured in the norm of $X^{(k)}$*, and are linearly dependent on* $\|f\|$*.*

Proof. (i) The proofs for Q and B are similar, so we only demonstrate it for Q. Let $p = (A, B, C, Q)$, $p^+ = (A, B, C, Q^+) \in D(S) \cap X^{(k)}$ and set $u := S(p)$ as well as $u^+ := S(p^+)$. By subtracting the equations that are solved by u and u^+ we conclude that $w := u^+ - u$ solves

$$(\mathcal{C}w')' + \mathcal{B}w' + (\mathcal{A} + \mathcal{Q})w = g_Q(u^+)[Q - Q^+]$$

in $L^2(I; V^*)$ and possesses homogeneous initial conditions. Additionally, the right-hand side lies in $H^k(I; H)$ and has a root of order $k-1$ at zero. Hence, the regularity results of Theorem 3.10 show that w fulfills the energy estimate

$$\begin{aligned}
\|w\|_{Y^{(k)}} &\leqslant \Lambda_Q \big\| g_Q(u^+)[Q - Q^+] \big\|_{H^k(I;H)} \\
&\leqslant \Lambda_Q \big\| u^+ \big\|_{H^k(I;V)} \big\| Q - Q^+ \big\|_{W^{k,\infty}(I;\mathcal{L}(V,H))} \\
&\leqslant \Lambda_Q \Lambda_{Q^+} \|f\| \big\| Q - Q^+ \big\|_{W^{k,\infty}(I;\mathcal{L}(V,H))},
\end{aligned}$$

where the positive constants Λ_Q and Λ_{Q^+} continuously depend on the $X^{(k)}$-norms of (A, B, C, Q) and (A, B, C, Q^+), respectively. The right-hand side f enters this inequality with its $H^k(I; H)$- or its $H^{k+1}(I; V^*)$-norm. The product $\Lambda_Q \Lambda_{Q^+} \|f\|$ is the Lipschitz constant.

(ii) Let (A, B, C, Q), $(A, B, C^+, Q) \in D(S) \cap X^{(k)}$. We use the same idea as in (i) for C, but have to be mindful of the initial conditions since they depend on C and C^+. However, due to $k \geqslant 1$ we have $u^+ \in H^2(I; H)$, i.e. $(u^+)'$ is continuous (taking values in H) and therefore $(\mathcal{C}(u^+)')(0)$ is well-defined and has to vanish. Hence,

$$(\mathcal{C}w')(0) = (\mathcal{C}(u^+)')(0) - (\mathcal{C}u')(0) = 0$$

holds in this case as well. The right-hand side of the evolution equation that is solved by w reads

$$g_C(u^+)[C - C^+] = ((\mathcal{C} - \mathcal{C}^+)(u^+)')',$$

which is an element of $H^{k-1}(H)$ and has vanishing derivatives at zero up to order $k-2$. Thus we can only apply the regularity theorem with $k-1$ instead of k, which yields

$$\begin{aligned}
\|w\|_{Y^{(k-1)}} &\leqslant \Lambda_C^{(k-1)} \big\| g_C(u^+)[C - C^+] \big\|_{H^{k-1}(I;H)} \\
&\leqslant \Lambda_C^{(k-1)} \big\| u^+ \big\|_{H^{k+1}(I;H)} \big\| C - C^+ \big\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(H))} \\
&\leqslant \Lambda_C^{(k-1)} \Lambda_{C^+}^{(k)} \|f\| \big\| C - C^+ \big\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(H))}.
\end{aligned}$$

The constant $\Lambda_C^{(k-1)}$ depends on (A, B, C, Q) in the $X^{(k-1)}$-norm, while $\Lambda_{C^+}^{(k)} > 0$ is influenced by the $X^{(k)}$-norm of (A, B, C^+, Q). This inequality is actually slightly stronger than the asserted Lipschitz continuity in $X^{(k)}$, because the difference of C and C^+ appears with its $W^{k,\infty}(I; \mathcal{L}^{sa}(H))$- and not its $W^{k+1,\infty}(I; \mathcal{L}^{sa}(H))$-norm. Unfortunately, we cannot benefit from this because $\Lambda_{C^+}^{(k)}$ is only well-defined if we assume C^+ to be $k+1$-times differentiable in time.

Estimates for A can be derived in the same fashion; in an analogous setting with A^+, the estimate turns out to be

$$\begin{aligned}
\|w\|_{Y^{(k-1)}} &\leqslant \Lambda_A^{(k-1)} \|g_A(u^+)[A - A^+]\|_{H^k(I;V^*)} \\
&\leqslant \Lambda_A^{(k-1)} \|u^+\|_{H^k(I;V)} \|A - A^+\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(V,V^*))} \\
&\leqslant \Lambda_A^{(k-1)} \Lambda_{A^+}^{(k)} \|f\| \|A - A^+\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(V,V^*))}.
\end{aligned}$$

Lipschitz continuity with respect to B and Q follows by weakening the assertion (i). We infer that $S\colon D(S) \cap X^{(k)} \to Y^{(k-1)}$ is locally Lipschitz continuous in all arguments with constants that also depend continuously on the other operators. Hence, the whole map is locally Lipschitz continuous as well. □

Now we can apply this theorem to show differentiability of S with respect to each of the four operators. We have seen in the previous lemma that the operators A and C must be treated differently from B and Q. This continues to be the case in the differentiability result.

Theorem 4.4. *Let $k \in \mathbb{N} \cup \{0\}$ and $f \in \mathcal{F}^{(k)}$. If $k \neq 0$ then we also assume $u_0 = u_1 = 0$. In this setting*

- *(i) the map $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$ is Fréchet-differentiable with respect to B and Q, and*
- *(ii) when $k \geqslant 2$, the map $S\colon D(S) \cap X^{(k)} \to Y^{(k-2)}$ is Fréchet-differentiable with respect to all arguments.*

In both cases, for each symbol $x \in \{A, B, C, Q\}$ and linearization argument h, the function $u_h = (\partial_x S)(A, B, C, Q)[h]$ is the unique solution of the equation

$$(\mathcal{C}u_h')' + \mathcal{B}u_h' + (\mathcal{A} + \mathcal{Q})u_h = g_x(u)[h]$$

in $L^2(I; V^*)$, under homogeneous initial conditions. Here, g_x is used as defined in (4.6), $p = (A, B, C, Q) \in D(S) \cap X^{(k)}$ and $u = S(p)$.

The space for h is the Banach space that the operators which are symbolized by x belong to. We only gave a rough definition of it in favor of a more pleasant formulation of the theorem.

Proof. (i) We prove the assertion for B; a very similar approach can be used for Q. Let $h \in W^{1,\infty}(I; \mathcal{L}(H))$ if $k = 0$ or $h \in W^{k,\infty}(I; \mathcal{L}(H))$ in the case $k > 0$. We define $u^+ := S(A, B + h, C, Q)$ and suppose that u_h is given as in the assertion. First, we note that the map $h \mapsto u_h$ is not merely linear but also bounded, as energy estimates for u_h make obvious. The difference $w \coloneqq u^+ - u - u_h$ solves the equation

$$(\mathcal{C}w')' + \mathcal{B}w' + (\mathcal{A} + \mathcal{Q})w = g_B(u - u^+)[h]$$

in $L^2(I; V^*)$ and has vanishing initial values. Both sides of this equation are regular enough to ensure that w fulfills the energy estimates of order k, obtained by Theorem 2.10 ($k = 0$) or Theorem 3.10 ($k > 0$). Together with Lemma 4.3 this gives rise to the calculation

$$\begin{aligned} \|w\|_{Y^{(k)}} &\leqslant \Lambda_B^{(k)} \big\| g_B(u - u^+)[h] \big\|_{H^k(I;H)} \\ &\leqslant \Lambda_B^{(k)} \big\| u - u^+ \big\|_{H^{k+1}(I;H)} \|h\|_{W^{k,\infty}(I;\mathcal{L}(H))} \\ &\leqslant \Lambda_B^{(k)} L_{B,B+h} \|h\|^2_{W^{k,\infty}(I;\mathcal{L}(H))}. \end{aligned} \tag{4.7}$$

Here, $L_{B,B+h}$ denotes the Lipschitz constant from the previous lemma. It continuously depends on h, thus it has to remain bounded as $h \to 0$. The constant $\Lambda_B^{(k)}$ does not depend on h. By combining these observations we conclude that $\|w\|_{Y^{(k)}} = \mathcal{O}\big(\|h\|^2\big)$ when we let h tend to zero in $W^{k,\infty}(I; \mathcal{L}(H))$. In particular, the map $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$ is differentiable with respect to B. Thus, we have shown that S possesses a partial Fréchet-derivative

$$\partial_B S\colon D(S) \cap X^{(k)} \to \mathcal{L}\Big(W^{k,\infty}(I; \mathcal{L}(H)), Y^{(k)}\Big).$$

(ii) We continue with A. The first difference to (i) is the fact that we cannot use all $h \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*))$. We have to make sure that h is small enough such that $p^+ \coloneqq (A + h, B, C, Q)$ belongs to $D(S) \cap X^{(k)}$ in order to make $u^+ \coloneqq S(p^+)$ well-defined. The set $D(S) \cap X^{(k)}$ is open in $X^{(k)}$, so this indeed only puts an upper bound on the norm of h. The local Lipschitz continuity of S and energy estimates (of order $k-2$) for $w = u^+ - u - u_h$ yield

$$\begin{aligned} \|w\|_{Y^{(k-2)}} &\leqslant \Lambda_A^{(k-2)} \big\| g_A(u - u^+)[h] \big\|_{H^{k-1}(I;V^*)} \\ &\leqslant \Lambda_A^{(k-2)} \big\| u - u^+ \big\|_{H^{k-1}(I;V)} \|h\|_{W^{k-1,\infty}(I;\mathcal{L}^{sa}(V,V^*))} \\ &\leqslant \Lambda_A^{(k-2)} L_{A,A+h} \|h\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(V,V^*))} \|h\|_{W^{k-1,\infty}(I;\mathcal{L}^{sa}(V,V^*))}. \end{aligned} \tag{4.8}$$

Although it looks like $\|w\| = \mathcal{O}\big(\|h\|^2\big)$ as $h \to 0$ in $W^{k,\infty}(I; \mathcal{L}^{sa}(V, V^*))$, this is incorrect because the Lipschitz constant $L_{A,A+h}$ is only bounded with respect to the $W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*))$-norm of h, hence we only obtain $\|w\| = \mathcal{O}\big(\|h\|^2\big)$ (and thus differentiability) within this smaller space.

The derivative of S with respect to C can be observed in a way that is similar to A, but we need to be mindful of the initial conditions. Let $h \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(H))$ be small and u^+ and w as before. Due to $k > 0$ we know that $u \in H^2(I; H)$, so not only $\mathcal{C}u'$ is continuous, but also u' itself, and it has the initial value $u'(0) = C(0)^{-1}(\mathcal{C}u')(0) = 0$. The same applies to u^+ and u_h, therefore w also fulfills homogeneous initial conditions in this case. The estimate for w reads

$$\begin{aligned}\|w\|_{Y^{(k-2)}} &\leqslant \Lambda_C^{(k-2)} \big\|g_C(u-u^+)[h]\big\|_{H^{k-2}(I;H)} \\ &\leqslant \Lambda_C^{(k-2)} \big\|u-u^+\big\|_{H^k(I;H)} \|h\|_{W^{k-2,\infty}(I;\mathcal{L}^{sa}(H))} \\ &\leqslant \Lambda_C^{(k-2)} L_{C,C+h} \|h\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(H))} \|h\|_{W^{k-1,\infty}(I;\mathcal{L}^{sa}(H))},\end{aligned} \tag{4.9}$$

which behaves like $\mathcal{O}\big(\|h\|^2\big)$ when $h \to 0$ in $W^{k+1,\infty}(I;\mathcal{L}^{sa}(H))$. □

We would like to remark again on the fact that although the derivative in direction A or C maps to $Y^{(k-1)}$, we can only show that it is indeed the derivative in the weaker norm of $Y^{(k-2)}$. In applications this is not an important restriction, because actual inverse problems will probably employ an L^2-type of space[2] instead of $Y^{(j)}$. However, it does imply that k has to be greater than or equal to two. This means that the unknowns A and C have to be three times weakly differentiable with respect to the time variable; in particular, they cannot have any discontinuities. For B and Q we can set k to zero, which means that we are able to differentiate S in operators B, Q that are once weakly differentiable in time.

4.2.1 *Validity of the tangential cone condition*

Before continuing with the Fréchet-derivative of the whole operator S, we would like to comment on the so-called *tangential cone condition* (TCC; also called *nonlinearity condition*).

Definition 4.5. We say that a Fréchet-differentiable operator $F\colon \mathcal{D}(F) \subset W \to Z$ between Banach spaces W and Z fulfills the tangential cone condition in $x^+ \in D(F)$ if there exist $r > 0$ and $0 \leqslant \omega < 1$ such that $B(x^+, r) \subset D(F)$ and

$$\big\|F(v) - F(w) - F'(w)[v-w]\big\|_Z \leqslant \omega \,\|F(v) - F(w)\|_Z \tag{4.10}$$

holds for all $v, w \in B(x^+, r)$.

From the Fréchet-differentiability we can only infer that the linearization error on the left-hand side is bounded by the W-norm of $v - w$, which might be significantly larger than the norms of $F(v) - F(w)$ or $F'(w)[v-w]$. With the TCC the linearization error can also be bounded by the difference of their function values, and the reverse triangle inequality shows $1 - \omega \leqslant \|F'(w)[v-w]\|/\|F(v) - F(w)\| \leqslant 1 + \omega$. This immediately provides equivalence of the local ill-posedness of F at x^+ and the local ill-posedness of $F'(x^+)$ (see [HS94] for a more general analysis). Moreover, if $F'(x^+)$ is injective, then $F(x) = F(x^+)$ implies $x = x^+$ for all $x \in B(x^+, r)$.

We see that the nonlinearity condition provides a strong connection between F and its derivative. Hence, it is not surprising that it is frequently used in the treatment of nonlinear inverse problems, for example

2 The measured data also has to belong to this image space, and observing differentiable measurement noise is probably not realistic.

to prove convergence results for regularization methods. However, in applications it is often not provable or even turns out to be false (especially for boundary data). For S we can give mixed results.

Corollary 4.6. *Let* $k \in \mathbb{N} \cup \{0\}$, $f \in \mathcal{F}^{(k)}$ *and* $p = (A, B, C, Q) \in D(S) \cap X^{(k)}$. *If* $k \neq 0$ *then we also assume* $u_0 = u_1 = 0$. *In this setting, both*

$$S(A, \cdot, C, Q)\colon W^{k_1,\infty}(I; \mathcal{L}(H)) \to Y^{(k)} \text{ and}$$
$$S(A, B, C, \cdot)\colon W^{k_1,\infty}(I; \mathcal{L}(V, H)) \to Y^{(k)}$$

satisfy the nonlinearity condition (4.10).

To avoid having to treat the cases $k = 0$ and $k > 0$ separately, we made use of the notation $k_1 := \max\{1, k\}$ for $k \in \mathbb{N} \cup \{0\}$.

Proof. For B we have already shown this in the proof of Theorem 4.4: For $F := S(A, \cdot, C, Q)$ equation (4.7) states

$$\begin{aligned}\big\|F(B+h) - F(B) - F'(B)[h]\big\|_{Y^{(k)}} \\ \leqslant \Lambda_B \|F(B+h) - F(B)\|_{H^{k+1}(I;H)} \|h\|_{W^{k,\infty}(I;\mathcal{L}(H))},\end{aligned}$$

and of course the set $Y^{(k)} = W^{k,\infty}(I; V) \cap W^{k+1,\infty}(I; H)$ is a subspace of $H^{k+1}(I; H)$. Further, the constant $\Lambda_B > 0$ continuously depends on B and is therefore bounded by $\Lambda > 0$ if we only regard operators B that belong to any fixed bounded subset of $W^{k_1,\infty}(I; \mathcal{L}(H))$. In such a subset we see that the nonlinearity condition (4.10) holds for all x^+, for example by choosing $\omega := 1/2$ and $r := 1/2\,\Lambda^{-1}$. The proof for $Q \mapsto S(A, B, C, Q)$ is obtained in the same way, and is very similar to the analysis in [GL17]. □

For A and C this is not possible: According to (4.8), we obtain for the operator $F := S(\cdot, B, C, Q)$ that the inequality

$$\begin{aligned}\big\|F(A+h) - F(A) - F'(A)[h]\big\|_{Y^{(k-2)}} \\ \leqslant \Lambda_A \|F(A+h) - F(A)\|_{H^{k-1}(I;V)} \|h\|_{W^{k+1,\infty}(I;\mathcal{L}(V,V^*))}\end{aligned}$$

holds for small h, but $\|F(A+h) - F(A)\|_{H^{k-1}(I;V)}$ can only be bounded by $\|F(A+h) - F(A)\|_{Y^{(k-1)}}$. Hence, we acquire something similar to (4.10), but the norms on both sides do not match. On the left-hand side we have $Y^{(k-2)}$, while the right-hand side has to be measured in the smaller space $Y^{(k-1)}$. The same effect can be observed for C through (4.9).

4.2.2 *Total differentiability*

For the reconstruction of one of the operators A, B, C and Q or a parameter that influences *exactly* one of these operators, the partial derivatives provided by Theorem 4.4 are sufficient. However, if the searched for quantity influences *multiple* operators then we also require the total derivative of S. We can infer its existence from Theorem 4.4 by proving that the partial derivatives of S are locally Lipschitz continuous. This fact is also interesting for the treatment of the inverse problems, because it allows to conclude ill-posedness of the derivative from ill-posedness of the nonlinear operator, see [HS94]. However, we will not make use of this technique because we can *directly* show ill-posedness for both S and its derivatives in the next section.

Lemma 4.7. *Let* $k \geqslant 2$, $f \in \mathcal{F}^{(k)}$ *and* $u_0 = u_1 = 0$. *Each of the operators*

$$\begin{aligned}
&\partial_A S \colon D(S) \cap X^{(k)} \to \mathcal{L}\Big(W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*)),\, Y^{(k-2)}\Big),\\
&\partial_B S \colon D(S) \cap X^{(k)} \to \mathcal{L}\Big(W^{k,\infty}(I; \mathcal{L}(H)),\, Y^{(k-1)}\Big),\\
&\partial_C S \colon D(S) \cap X^{(k)} \to \mathcal{L}\Big(W^{k+1,\infty}(I; \mathcal{L}^{sa}(H)),\, Y^{(k-2)}\Big) \text{ and}\\
&\partial_Q S \colon D(S) \cap X^{(k)} \to \mathcal{L}\Big(W^{k,\infty}(I; \mathcal{L}(V, H)),\, Y^{(k-1)}\Big)
\end{aligned}$$

is locally Lipschitz continuous.

Proof. The proofs mainly differ in the use of different spaces, and the most difficult cases to treat are $\partial_A S$ and $\partial_C S$. Therefore we only present a proof for the partial derivative of S with respect to C.

Let $h \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(H))$ and for $i = 1, 2$ let $p_i = (A_i, B_i, C_i, Q_i) \in D(S) \cap X^{(k)}$ be given. Moreover, we set $u_{h,i} := \partial_C S(p_i)[h]$ and abuse notation by defining $u_i := S(p_i)$. Our task is to find $L > 0$ such that $\|u_{h,2} - u_{h,1}\|$ is bounded by $L\|p_2 - p_1\|\|h\|$. Contrary to previous proofs, the weak formulations of $u_{h,i}$ differ in their left- *and* right-hand sides. In order to connect them, we introduce the function w_h which solves the equation with the left-hand side of $u_{h,1}$ and the right-hand side of $u_{h,2}$. Thus, in addition to homogeneous initial conditions, w_h is supposed to satisfy

$$(\mathcal{C}_1 w_h')' + \mathcal{B}_1 w_h' + (\mathcal{A}_1 + \mathcal{Q}_1) w_h = g_C(u_2)[h]$$

in the $L^2(I; V^*)$-sense. Since $u_{h,1}$ and w_h solve the same formulation with different right-hand sides, Theorem 3.10 provides an energy estimate for $u_{h,1} - w_h$. This estimate is of the form

$$\begin{aligned}
\|u_{h,1} - w_h\|_{Y^{(k-2)}} &\leqslant \Lambda_1 \|g_C(u_1 - u_2)[h]\|_{H^{k-2}(I;H)}\\
&\leqslant \Lambda_1 \|g_C\| \|u_1 - u_2\|_{H^k(I;H)} \|h\|_{W^{k-1,\infty}(I;\mathcal{L}^{sa}(H))},
\end{aligned} \tag{4.11}$$

with a constant Λ_1 that continuously depends on the $X^{(k-2)}$-norm of p_1. With $\|g_C\|$ we denote the norm of g_C in the space

$$\mathcal{L}\big(H^k(I;H),\, \mathcal{L}\big(W^{k-1,\infty}(I; \mathcal{L}(H)),\, H^{k-2}(I;H)\big)\big),$$

thus $\|g_C\|$ only depends on k. For the estimation of $\|u_1 - u_2\|$ in (4.11) we employ the local Lipschitz continuity of $S \colon D(S) \cap X^{(k)} \to Y^{(k-1)}$ (see Lemma 4.3) to obtain another constant Λ_2 depending on p_1 in the space $X^{(k)}$, as well as the estimate

$$\|u_{h,1} - w_h\|_{Y^{(k-2)}} \leqslant \Lambda_1 \Lambda_2 \|g_C\| \|p_1 - p_2\|_{X^{(k)}} \|h\|_{W^{k-1,\infty}(I;\mathcal{L}^{sa}(H))}. \tag{4.12}$$

Next, we turn to the distance between $u_{h,2}$ and w_h. By design of w_h, these functions solve an evolution equation with the same right-hand side, but different left-hand sides. Hence, we can apply Lipschitz continuity of the operator S, which arises when the right-hand side f is replaced by $g_C(u_2)[h] \in H^{k-2}(I;H)$. Due to linearity of the equation, the norm of the right-hand side has to enter linearly into the Lipschitz constant, thus (again using Lemma 4.3) we gain a constant $\Lambda_3 > 0$ depending continuously on p_1 in $X^{(k-1)}$, such that

$$\begin{aligned} \|u_{h,2} - w_h\|_{Y^{(k-2)}} &\leqslant \Lambda_3 \|g_C(u_2)[h]\|_{H^{k-1}(I;H)} \|p_1 - p_2\|_{X^{(k-1)}} \\ &\leqslant \Lambda_3 \|g_C\| \|u_2\|_{H^{k+1}(I;H)} \|h\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(H))} \|p_1 - p_2\|_{X^{(k-1)}}. \end{aligned} \tag{4.13}$$

This time, $\|g_C\|$ denotes the norm of g_C in the space

$$\mathcal{L}\big(H^{k+1}(I;H),\ \mathcal{L}\big(W^{k,\infty}(I;\mathcal{L}^{sa}(H)),\ H^{k-1}(I;H)\big)\big).$$

We can eliminate the dependence of (4.13) on u_2 by using the regularity estimates for $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$. They provide $\Lambda_4 > 0$ with

$$\|u_{h,2} - w_h\|_{Y^{(k-2)}} \leqslant \Lambda_3 \Lambda_4 \|g_C\| \|p_1 - p_2\|_{X^{(k-1)}} \|f\| \|h\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(H))},$$

where f is measured as usual, using the $H^k(I;H)$- or $H^{k+1}(I;V^*)$-norm.

Finally, we can combine the last inequality with (4.12) to conclude

$$\begin{aligned} \|u_{h,1} - u_{h,2}\|_{Y^{(k-2)}} &\leqslant \|u_{h,1} - w_h\|_{Y^{(k-2)}} + \|u_{h,2} - w_h\|_{Y^{(k-2)}} \\ &\leqslant L \|f\| \|p_1 - p_2\|_{X^{(k-1)}} \|h\|_{W^{k,\infty}(I;\mathcal{L}^{sa}(H))} \\ &\leqslant L \|f\| \|p_1 - p_2\|_{X^{(k-1)}} \|h\|_{W^{k+1,\infty}(I;\mathcal{L}^{sa}(H))}. \end{aligned}$$

The constant $L > 0$ is comprised of $\Lambda_1, \dots, \Lambda_4$ and thus continuously depends on p_1 and p_2 in $X^{(k)}$. □

The total differentiability of the operator S immediately follows from its differentiability in all arguments and the continuity of the derivatives.

Corollary 4.8. *Let $k \geqslant 2$, $u_0 = u_1 = 0$ and $f \in \mathcal{F}^{(k)}$. The operator*

$$S\colon D(S) \cap X^{(k)} \subset X^{(k)} \to Y^{(k-2)}$$

is Fréchet-differentiable in every $p = (A, B, C, Q) \in D(S) \cap X^{(k)}$. The value $\partial S(p)[h]$ of its derivative at $h = (\bar{A}, \bar{B}, \bar{C}, \bar{Q}) \in X^{(k)}$ is given as the solution $u_h \in Y^{(k-1)}$ of the evolution equation

$$\begin{aligned} &(\mathcal{C}u_h')' + \mathcal{B}u_h' + (\mathcal{A} + \mathcal{Q})u_h \\ &\qquad = g_A(u)\big[\bar{A}\big] + g_B(u)\big[\bar{B}\big] + g_C(u)\big[\bar{C}\big] + g_Q(u)\big[\bar{Q}\big] \\ &\qquad = -\big(\bar{C}(\cdot)[u'(\cdot)]\big)' - \bar{B}(\cdot)[u'(\cdot)] - \big(\bar{A}(\cdot) + \bar{Q}(\cdot)\big)[u(\cdot)], \end{aligned}$$

equipped with homogeneous initial conditions $u_h(0) = (\mathcal{C}u_h')(0) = 0$. Moreover, the derivative

$$\partial S\colon D(S) \cap X^{(k)} \subset X^{(k)} \to \mathcal{L}\Big(X^{(k)},\ Y^{(k-2)}\Big)$$

is locally Lipschitz continuous. As always, $u = S(p) \in Y^{(k)}$ denotes the solution of the forward problem.

One of the main conclusions we can draw from this result is that S maps to $Y^{(k)}$ and its derivative to $Y^{(k-1)}$, but differentiability only holds with respect to $Y^{(k-2)}$.

4.2.3 *Adjoint of the derivative*

For the numerical inversion of the linearized problems that arise from S, we require not only its Fréchet-derivative, but further need to be able to evaluate the adjoint of this derivative. However, at this point we only know that this adjoint has to exist, but have no means of calculating it efficiently.

Of course the adjoint is hugely dependent on the choice of spaces for the domain and co-domain of S. From an application viewpoint, $Y^{(k)}$ is not a suitable space for measured data, because this would imply that the noise introduced by sensors is $k+1$ times differentiable in time. Furthermore, the spatial regularity that we attribute to V does not seem appropriate as well. Thus, an approach with L^2-spaces seems more sensible here, and also makes the analysis easier because $L^2(I;H)$ is a Hilbert space. We will however not reduce the size of the domain of S, mainly because the abstract operator setting does not allow any "easy" spaces here. For example, there is no way of transforming $\mathcal{L}(H)$ into a Hilbert space and simultaneously not restricting possible applications too much.[3]

Therefore we seek to calculate the adjoint of $\partial S(p) \in \mathcal{L}(X^{(k)}, L^2(I;H))$, which can be identified with an operator $\partial S(p)^* \in \mathcal{L}\big(L^2(I;H), \big(X^{(k)}\big)^*\big)$. But even for this choice of spaces, the application of $\partial S(p)^*[v] \in \big(X^{(k)}\big)^*$ to $h \in X^{(k)}$ must still be calculated by $(v, \partial S(p)h)_{L^2(I;H)}$ and therefore involves the solution of a different differential equation for every h. Hence, we will try to shift as many operations from h to v as possible.

Unsurprisingly, the efficient evaluation of $\partial S(p)^*[v]$ will involve the solution of another evolution equation, namely of the adjoint equation to (4.1a), which is

$$(\mathcal{C}w')' - \mathcal{B}^*w' + (\mathcal{A} + \mathcal{Q}^* - (\mathcal{B}^*)')w = v \text{ in } L^2(I;V^*) \tag{4.14a}$$

and has to be furnished with homogeneous *end conditions*

$$w(T) = (\mathcal{C}w')(T) = 0. \tag{4.14b}$$

Here, $\mathcal{B}^*$ and $\mathcal{Q}^*$ denote the realizations of $t \mapsto B(t)^*$ and $t \mapsto Q(t)^*$ respectively. However, due to their pointwise definition they coincide with the adjoints of $\mathcal{B} \in \mathcal{L}(L^2(I;H))$ and $\mathcal{Q} \in \mathcal{L}(L^2(I;V), L^2(I;H))$.

If $B = 0$ and Q is pointwise self-adjoint, then this is the original equation which has to be solved backwards in time. Otherwise (4.14a) has a structural flaw, which we need to address before we can solve it with the results of the preceding chapters: The operator $Q(t) \in \mathcal{L}(V, H)$ is only guaranteed to possess an adjoint $Q(t)^* \in \mathcal{L}(H, V^*)$, which does not

3 Even for the treatment of the PDEs in Chapters 5 and 6 this will turn out to be impractical.

fit into our framework. Only if we make the assumption $Q(t)^* \in \mathcal{L}(V, H)$ we are able to solve (4.14) using Theorem 2.10. The end conditions become initial conditions after reversing time with the transformation $t \mapsto T - t$. Apart from that, this change in coordinates only has the effect of removing the negative sign in the term $-\mathcal{B}^* w'$.

The operator $v \mapsto w$ that arises from (4.14) is the adjoint of the operator $(f \mapsto u) \in \mathcal{L}(L^2(I; H))$, where f is the right-hand side of the original evolution problem (4.1) that is solved by u. The fact that the derivative itself also involves the solution of this evolution equation already motivates why its adjoint might be connected to (4.14). The following theorem takes care of the details. In its formulation we make use of $k_1 := \max\{1, k\}$ once more, and by abuse of notation write Q^* although we actually mean $t \mapsto Q(t)^*$.

Theorem 4.9. *Let* $p = (A, B, C, Q) \in D(S) \cap X^{(k)}$ *with* $Q^* \in L^\infty(I; \mathcal{L}(V, H))$, $k \in \mathbb{N} \cup \{0\}$, $f \in \mathcal{F}^{(k)}$ *and* $u = S(p)$. *In the case that* $k \geqslant 1$ *we further assume* $u_0 = u_1 = 0$. *For* $v \in L^2(I; H)$ *let* $w_v \in Y$ *denote the unique solution of problem* (4.14).

(i) The adjoints of

$$\partial_Q S(p) \in \mathcal{L}\big(W^{k_1,\infty}(I; \mathcal{L}(V, H)),\ L^2(I; H)\big) \quad \text{and}$$
$$\partial_B S(p) \in \mathcal{L}\big(W^{k_1,\infty}(I; \mathcal{L}(H)),\ L^2(I; H)\big)$$

can for all $\bar{Q} \in W^{k_1,\infty}(I; \mathcal{L}(V, H))$ *and* $\bar{B} \in W^{k_1,\infty}(I; \mathcal{L}(H))$ *be characterized via*

$$\big\langle (\partial_Q S(p))^*[v], \bar{Q} \big\rangle = -\int_0^T \big(\bar{Q}(t)u(t), w_v(t)\big)\,dt \quad \text{and}$$
$$\big\langle (\partial_B S(p))^*[v], \bar{B} \big\rangle = -\int_0^T \big(\bar{B}(t)u'(t), w_v(t)\big)\,dt.$$

(ii) If $k \geqslant 2$, *then the evaluation of the adjoints of*

$$\partial_A S(p) \in \mathcal{L}\big(W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*)),\ L^2(I; H)\big) \quad \text{and}$$
$$\partial_C S(p) \in \mathcal{L}\big(W^{k+1,\infty}(I; \mathcal{L}^{sa}(H)),\ L^2(I; H)\big)$$

can for every $\bar{A} \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*))$ *and* $\bar{C} \in W^{k+1,\infty}(I; \mathcal{L}^{sa}(H))$ *be expressed through*

$$\big\langle (\partial_A S(p))^*[v], \bar{A} \big\rangle = -\int_0^T \big\langle \bar{A}(t)u(t), w_v(t) \big\rangle\,dt \quad \text{and}$$
$$\big\langle (\partial_C S(p))^*[v], \bar{C} \big\rangle = \int_0^T \big(\bar{C}(t)u'(t), w_v'(t)\big)\,dt.$$

Proof. Theorem 2.10 states that w_v is well-defined for all $v \in L^2(I; H)$ and that $v \mapsto w$ belongs to $\mathcal{L}(L^2(I; H), Y)$. From Theorem 4.4 we know that $\partial_x S(p)[h] = u_h$ holds for each symbol $x \in \{A, B, C, Q\}$, where u_h solves

$$(\mathcal{C}u_h')' + \mathcal{B}u_h' + (\mathcal{A} + \mathcal{Q})u_h = g_x(u)[h]$$

with homogeneous initial conditions and g_x is as defined in (4.6). We test equation (4.14a) that is solved by w_v at time $t \in I$ with $u_h(t)$ and integrate over the time interval I to see

$$
\begin{aligned}
(v, u_h)_{L^2(I;H)} = \int_0^T & \left\langle (\mathcal{C}w_v')'(t), u_h(t) \right\rangle + \langle A(t) w_v(t), u_h(t) \rangle \\
& - \left(B^*(t) w_v'(t), u_h(t)\right) + \left(Q^*(t) w_v(t), u_h(t)\right) \\
& - \left((B^*)'(t) w_v(t), u_h(t)\right) \mathrm{d}t.
\end{aligned}
\tag{4.15}
$$

On the first expression in the integral on the right-hand side of (4.15) we apply the integration by parts formula provided by Theorem A.11 twice, once in $V \subset H \subset V^*$ and once in $H \subset V^* \subset H^*$. The latter is a Gelfand triple as well, and is obtained by identifying V^* with V through its inner product (and keeping H and H^* separate). We conclude

$$
\begin{aligned}
\int_0^T \left\langle (\mathcal{C}w_v')'(t), u_h(t) \right\rangle \mathrm{d}t &= -\int_0^T \left((\mathcal{C}w_v')(t), u_h'(t)\right) \mathrm{d}t \\
&= \int_0^T \left((\mathcal{C}u_h')'(t), w_v(t)\right) \mathrm{d}t
\end{aligned}
\tag{4.16}
$$

because $C(t)$ is self-adjoint. The boundary terms vanish due to the homogeneous initial and end conditions of u_h and w_v, respectively. For expressions in (4.15) that involve B, we can use the product rule (Lemma 2.3) to see

$$
\begin{aligned}
&\int_0^T \left(B^*(t) w_v'(t), u_h(t)\right) + \left((B^*)'(t) w_v(t), u_h(t)\right) \mathrm{d}t \\
&\quad = \int_0^T \left((\mathcal{B}^* w_v)'(t), u_h(t)\right) \mathrm{d}t = -\int_0^T \left(B(t) u_h'(t), w_v(t)\right) \mathrm{d}t.
\end{aligned}
\tag{4.17}
$$

Dealing with Q and A is simple because we only need to insert their adjoints, bearing in mind that $A(t)$ is self-adjoint. Now we plug (4.16) and (4.17) into (4.15) to obtain

$$
\begin{aligned}
(v, u_h)_{L^2(I;H)} = \int_0^T & \left\langle (\mathcal{C}u_h')'(t), w_v(t) \right\rangle + \left(B(t) u_h'(t), w_v(t)\right) \\
& + \langle A(t) u_h(t), w_v(t) \rangle + \left(Q(t) u_h(t), w_v(t)\right) \mathrm{d}t,
\end{aligned}
$$

which contains the left-hand side of the equation that is solved by $u_h(t)$, tested with $w_v(t)$. We replace it by the corresponding right-hand side and arrive at

$$
\begin{aligned}
\langle \partial_x S(p)^*[v], h \rangle &= (v, \partial_x S(p)[h])_{L^2(I;H)} \\
&= (v, u_h)_{L^2(I;H)} = \int_0^T \langle g_x(u)[h](t), w_v(t) \rangle \,\mathrm{d}t.
\end{aligned}
$$

The assertion follows by inserting the definition of g_x. In the case of $\partial_C S(p)^*$, we can use the integration by parts formula one additional time to get rid of the time derivative acting on h. □

With this result the application of $\partial S(p)^*[v]$ on $h \in X^{(k)}$ can be implemented efficiently, because the effort of computing w_v does not depend on h. Furthermore, the operations that do depend on h (multiplication, integration over I) are cheap. Unfortunately, we are not able to represent the adjoint completely by means of the inner product of $L^2(I;H)$, since we do not know how e.g. the application of $\bar{C}(t)$ on $u'(t)$ looks like. Hence, we cannot shift $u'(t)$ to $w_v(t)$. However, this will be the case when we apply this theory to actual partial differential equations, because then we have more information about the structure of the operators.

As a direct consequence of the above theorem, we can also describe the adjoint of $\partial S(p)$.

Corollary 4.10. *Let the assumptions of Theorem 4.9 be fulfilled with* $k \geqslant 2$. *The adjoint* $(\partial S(p))^* \in \mathcal{L}\big(L^2(I;H), (X^{(k)})^*\big)$ *of* $\partial S(p) \in \mathcal{L}\big(X^{(k)}, L^2(I;H)\big)$ *can be evaluated at* $v \in L^2(I;H)$ *and* $h = (\bar{A}, \bar{B}, \bar{C}, \bar{Q}) \in X^{(k)}$ *by*

$$\langle (\partial S(p))^*[v], h\rangle_{(X^{(k)})^* \times X^{(k)}} = \int_0^T \big(\bar{C}(t)u'(t), w_v'(t)\big) - \big(\bar{B}(t)u'(t), w_v(t)\big) - \big\langle \big(\bar{A}(t) + \bar{Q}(t)\big)u(t), w_v(t)\big\rangle \,\mathrm{d}t.$$

We wish to close this section with the remark that changing the codomain of S from Y to $L^2(I;H)$ was not only important for possible applications. Without this modification, we would not have been able to solve the adjoint equation (4.14a), because then the right-hand side v would only have been an element of Y^*. This is a superset of $L^2(I;V^*)$ and, as we know from the preceding chapters, is therefore not sufficiently regular to provide a solution w_v.

4.3 ILL-POSEDNESS

In particular for the numerical treatment of inverse problems, it is crucial to know whether the task under consideration is ill-posed or well-posed, because this fact determines the choice of algorithms. With regards to Newton-based solvers, the Fréchet-derivative of the parameter-to-state map is also of natural interest. Therefore, we will discuss the ill-posedness of S and its linearization in this section.

We make use of the usual concepts of ill-posedness of linear- and nonlinear operators. In particular, a linear operator is ill-posed if and only if its generalized inverse is discontinuous. A nonlinear operator G is locally ill-posed at a point x if and only if in every neighborhood of x there exists a sequence that does not converge to x, although the corresponding images do converge to $G(x)$. For more details we refer to the literature, e.g. [Rie03].

For Fréchet-differentiability we can employ the chain rule to see that $F = \Psi \circ S \circ P$ (as depicted in Figure 4.1 on page 44) is differentiable if all of the involved operators are differentiable, which is something we can expect in applications. Naturally, local ill-posedness of S transfers to $\Psi \circ S$ if the measurement operator Ψ is at least continuous. By the same argument, ill-posedness of F is easily proven if the operator P, that maps searched for parameters to the operators A, B, C and Q, is ill-posed. However, we do not want to base our analysis on the unknown operator P. In particular, the image of the operator P might not contain any point at which S is locally ill-posed. Even if it does, it is not evident that the perturbations which can be used to show ill-posedness of S also lie in the image of P and have nonconvergent preimages.[4] Therefore we do not prove the local ill-posedness of S directly. Instead, we first give an intermediate result that can also be used to show ill-posedness in a setting where not the operators themselves, but another parameter that influences them is sought. For both situations we need to be aware in which circumstances the images of a sequence of parameters under S converge.

Theorem 4.11. *Let $k \in \mathbb{N} \cup \{0\}$, $\delta > 0$ and $f \in \mathcal{F}^{(k)}$. If $k \geqslant 1$ then we also require $u_0 = u_1 = 0$. Further, let $p = (A, B, C, Q) \in D(S) \cap X^{(k)}$, $u := S(p)$ and $k_1 := \max\{1, k\}$. Then the following holds:*

(i) If $(R_j)_{j\in\mathbb{N}} \subset W^{k_1,\infty}(I; \mathcal{L}(V, H))$ satisfies $\|R_j\| \leqslant \delta$ and

$$\mathcal{R}_j v \to 0 \quad \textit{in } H^k(I; H) \textit{ for all } v \in Y^{(k)},$$

then $S(A, B, C, Q + R_j) \to u$ in $Y^{(k)}$ as $j \to \infty$.

(ii) If $(R_j)_{j\in\mathbb{N}} \subset W^{k_1,\infty}(I; \mathcal{L}(H))$ satisfies $\|R_j\| \leqslant \delta$ and

$$\mathcal{R}_j v' \to 0 \quad \textit{in } H^k(I; H) \textit{ for all } v \in Y^{(k)},$$

then $S(A, B + R_j, C, Q) \to u$ in $Y^{(k)}$ as $j \to \infty$.

Suppose additionally that $k > 0$. Then we also claim the following.

(iii) Let $(R_j)_{j\in\mathbb{N}} \subset W^{k+1,\infty}(I; \mathcal{L}(V, V^*))$ with $\|R_j\| \leqslant \delta$, and suppose that δ is small enough to ensure $(A + R_j, B, C, Q) \in D(S)$ for all $j \in \mathbb{N}$, and

$$\mathcal{R}_j v \to 0 \quad \textit{in } H^k(I; V^*) \textit{ for all } v \in Y^{(k)}.$$

Then $S(A + R_j, B, C, Q) \to u$ in $Y^{(k-1)}$ when $j \to \infty$.

(iv) Let $(R_j)_{j\in\mathbb{N}} \subset W^{k+1,\infty}(I; \mathcal{L}(H))$ with $\|R_j\| \leqslant \delta$, and suppose that δ is small enough to ensure $(A, B, C + R_j, Q) \in D(S)$ for all $j \in \mathbb{N}$, and

$$(\mathcal{R}_j v')' \to 0 \quad \textit{in } H^{k-1}(I; H) \textit{ for all } v \in Y^{(k)}.$$

Then $S(A, B, C + R_j, Q) \to u$ in $Y^{(k-1)}$ when $j \to \infty$.

In each case the convergence is uniform in (A, B, C, Q) on every bounded subset of $D(S) \cap X^{(k)}$.

4 Example: $S \colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x\mathbb{1}_{\mathbb{R}_{\geqslant 0}}(x)$ is locally ill-posed at 0. However, $S \circ P \colon \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ with $P := \mathrm{Id}_{\mathbb{R}_{\geqslant 0}}$ is well-posed, although $0 \in P(\mathbb{R}_{\geqslant 0})$.

The uniform convergence becomes important when a searched for quantity influences not only one, but multiple operators. Also note that one could have achieved a more compact formulation of this theorem by using $g_x(v)[R_j]$. We have not done so in order to emphasize the different function spaces that are involved.

Proof. (i) We start with Q. Let $u_j = S(A, B, C, Q + R_j)$. The fields u and u_j solve

$$\begin{aligned} (\mathcal{C}u')' + \mathcal{B}u' + (\mathcal{Q} + \mathcal{R}_j)u + \mathcal{A}u &= f + \mathcal{R}_j u \quad \text{and} \\ (\mathcal{C}u_j')' + \mathcal{B}u_j' + (\mathcal{Q} + \mathcal{R}_j)u_j + \mathcal{A}u_j &= f, \end{aligned}$$

respectively, and have the same initial values. Hence, $w_j = u - u_j$ is a solution to

$$(\mathcal{C}w_j')' + \mathcal{B}w_j' + (\mathcal{Q} + \mathcal{R}_j)w_j + \mathcal{A}w_j = \mathcal{R}_j u$$

with homogeneous initial conditions and therefore satisfies the energy estimate

$$\|w_j\|^2_{Y(k)} \leqslant \Lambda_j^2 \|\mathcal{R}_j u\|^2_{H^k(I;H)} \tag{4.18}$$

with $\Lambda_j > 0$. The constants in the energy estimates continuously depend on the operators in the evolution equation (cf. Theorem 3.10), and the sequence $(R_j)_{j\in\mathbb{N}}$ of perturbations is bounded. Thus, the sequence $(\Lambda_j)_{j\in\mathbb{N}}$ is bounded as well. Together with the properties of R_j this implies $\|w_j\| \to 0$ as $j \to \infty$. This convergence is uniform in A, B, C and Q because both Λ_j ($j \in \mathbb{N}$) and u depend continuously on them.

(ii) The proof for B can be done in the same fashion, instead of (4.18) we obtain

$$\|w_j\|^2_{Y(k)} \leqslant \lambda_j^2 \|\mathcal{R}_j u'\|^2_{H^k(I;H)} \to 0.$$

(iii) For the other two operators we lose one order of regularity, because the right-hand side of the equation that is solved by $w_j = u - u_j$ is less regular. When we perturb A, the w_j satisfy

$$\|w_j\|^2_{Y(k-1)} \leqslant \lambda_j^2 \|\mathcal{R}_j u\|^2_{H^k(I;V^*)},$$

which vanishes in the limit $j \to \infty$.

(iv) In the case of C, the estimate reads

$$\|w_j\|^2_{Y(k-1)} \leqslant \lambda_j^2 \left\|(\mathcal{R}_j u')'\right\|^2_{H^{k-1}(I;H)} \to 0. \qquad \square$$

Even in this general framework, such sequences R_j always exist; we do not even have to use the time variable. The following lemma provides the basic building blocks for their construction.

Lemma 4.12. *There exist constants $\Gamma > \gamma > 0$ and sequences of operators*

(i) $(X_j)_{j\in\mathbb{N}} \subset \mathcal{L}^{sa}(H)$ such that $X_j v \to 0$ in H as $j \to \infty$ for all fixed $v \in H$ and $\Gamma \geqslant \|X_j\| \geqslant \gamma$ in $\mathcal{L}^{sa}(H)$ and $\mathcal{L}^{sa}(V, V^*)$, and

(ii) $(Y_j)_{j\in\mathbb{N}} \subset \mathcal{L}(V)$ *with* $Y_j v \to 0$ *in* V *as* $j \to \infty$ *for all fixed* $v \in V$ *and* $\Gamma \geqslant \|Y_j\| \geqslant \gamma$ *in* $\mathcal{L}(V)$ *and* $\mathcal{L}(V, H)$.

Proof. From the pointwise convergence (and therefore boundedness) of the operators we can already deduce the existence of the upper bound Γ using the uniform boundedness principle.

(i) Let $(\varphi_j)_{j\in\mathbb{N}} \subset V$ denote an orthonormal basis of H (possible because V is dense in H). We use it to define X_j for $v \in H$ as

$$X_j v := (v, \varphi_j)_H\, \varphi_j. \tag{4.19}$$

Apparently $\|X_j v\|_H = |(v, \varphi_j)_H| \to 0$ for $v \in H$ and $\|X_j\|_{\mathcal{L}(H)} \leqslant 1$. By evaluating X_j at φ_j we can also see $\|X_j\| \geqslant 1$. For $u, v \in V$ we have

$$\langle X_j v, u\rangle_{V^*\times V} = (v, \varphi_j)_H \langle \varphi_j, u\rangle_{V^*\times V} = (v, \varphi_j)_H (u, \varphi_j)_H,$$

which implies $\|X_j\|_{\mathcal{L}(V,V^*)} \leqslant 1$ and due to $\varphi_j \in V$ we may set $u = v = \varphi_j$ to infer $\|X_j\|_{\mathcal{L}(V,V^*)} \geqslant 1$. We have already shown that $X_j v \to 0$ in H for $v \in H$, thus $X_j v \to 0$ in V^* for all $v \in V$.

(ii) For $\mathcal{L}(V)$ we could use the same X_j if we replace φ_j by an orthonormal basis $(\psi_j)_{j\in\mathbb{N}}$ of V, but this sequence would not be suitable for $\mathcal{L}(V, H)$ due to compactness of V in H (orthonormal bases of V convergence strongly to zero in H and V^*). Hence, we modify equation (4.19) slightly to arrive at

$$Y_j v := (v, \psi_j)_V\, \psi_1,$$

which works in this case because $Y_j v \to 0$ for $v \in V$, but at the expense that Y_j is not self-adjoint. □

Finally, we can show local ill-posedness of S, even with data in $Y^{(k)}$. Of course this also proves the ill-posedness for data belonging to $L^2(I; H)$ because convergence in $Y^{(k)}$ implies convergence in $L^2(I; H)$.

Theorem 4.13. *Let* $k \in \mathbb{N} \cup \{0\}$ *and* $f \in \mathcal{F}^{(k)}$. *If* $k \geqslant 1$ *then we assume* $u_0 = u_1 = 0$. *Let* $(A, B, C, Q) \in D(S) \cap X^{(k)}$.

(i) The tasks of finding B *or* Q *such that* $S(A, B, C, Q) = y \in Y^{(k)}$ *holds are locally ill-posed in* B *and* Q.

(ii) Suppose $k \geqslant 1$. *Then the problems of finding operators* A *or* C *such that* $S(A, B, C, Q) = y \in Y^{(k-1)}$ *holds are locally ill-posed in* A *and* C.

Proof. We prove the claim by explicitly constructing sequences of operators that do not converge, but stay arbitrary close to $p = (A, B, C, Q)$ such that their image under S converges to $S(p)$. To this end, fix some $r > 0$.

(i) We start with Q and set $Q_j(t) := Q(t) + \tilde{r}\, Y_j$ with $\tilde{r} := r/\Gamma$ and $\Gamma > 0$, $Y_j \in \mathcal{L}(V, H)$ as in Lemma 4.12. This way $Q_j \in B(Q, r)$ and $Q_j \not\to Q$ in $W^{k,\infty}(I; \mathcal{L}(V, H))$. We show that $R_j(t) := \tilde{r}\, Y_j$ $(j \in \mathbb{N})$ satisfy the requirements of case (i) in Theorem 4.11 with $\delta := r$. For $v \in Y^{(k)}$ we have

$$\left\|\tilde{r}\, Y_j v\right\|^2_{H^k(I;H)} = \tilde{r}^2 \sum_{i=0}^{k} \int_0^T \left\|Y_j v^{(i)}(t)\right\|^2_H \,dt. \tag{4.20}$$

Every one of the finitely many integrands converges pointwise to zero and is bounded by $\Gamma^2 \|v^{(i)}(t)\|^2_V \in L^1(I)$, hence the whole sum vanishes in the limit $j \to \infty$.

For B we have to use $(X_j)_{j\in\mathbb{N}} \subset \mathcal{L}(H)$ as in Lemma 4.12 as the perturbation. Instead of (4.20) we obtain

$$\left\|\tilde{r}\, X_j v'\right\|^2_{H^k(I;H)} = \tilde{r}^2 \sum_{i=0}^{k} \int_0^T \left\|X_j v^{(i+1)}(t)\right\|^2_H \,dt,$$

which converges to zero for similar reasons. The convergence of $S(p_j)$ to $S(p)$ then follows from assertion (ii) in Theorem 4.11.

(ii) We set $A_j(t) = A(t) + \tilde{r}\, X_j$ and $C_j(t) = C(t) + \tilde{r}\, X_j$, again using $\tilde{r} := r/\Gamma$. Since $D(S)$ is open, the resulting p_j belong to $D(S)$ as long as r is sufficiently small. For every $v \in Y^{(k-1)}$ we have

$$\left\|\tilde{r}\, X_j v\right\|^2_{H^k(I;V^*)} = \tilde{r}^2 \sum_{i=0}^{k} \int_0^T \left\|X_j v^{(i)}(t)\right\|^2_{V^*} \,dt,$$

which converges to zero in the limit.

For C the reasoning is similar, we obtain

$$\left\|(\tilde{r}\, X_j v')'\right\|^2_{H^k(I;H)} = \tilde{r}^2 \sum_{i=0}^{k} \int_0^T \left\|X_j v^{(i+2)}(t)\right\|^2_H \,dt$$

and use $v \in H^{k+2}(I;H)$. In both cases we can apply Theorem 4.11, again with $\delta := r$. □

For convenience we used sequences of perturbations that are time-independent, which confirms that the corresponding "static" problems are ill-posed as well. In the case of time-dependent functions we would have to ensure that they are smooth enough to belong to $X^{(k)}$, which requires some work. However, it is also possible to use functions that only depend on time, so the ill-posedness is not due to the spatial variable. We will showcase this in Section 5.4, where we apply Theorem 4.11 to the acoustic wave equation.

The ill-posedness of the linearized problem can be concluded from the local ill-posedness of S, because we showed its (local) Lipschitz continuity in Lemma 4.3. However, we can also show it directly through compact embeddings, which are established in the following lemma.

Lemma 4.14 (Aubin-Lions, 1963). *Let X_0, X and X_1 be Banach spaces with $X_0 \subset X \subset X_1$, where the embedding $X_0 \hookrightarrow X$ is presumed to be compact and $X \hookrightarrow X_1$ to be continuous. For $1 \leqslant p, q \leqslant \infty$ let*

$$W_{p,q}(I; X_0, X_1) := \left\{\, u \in L^p(I; X_0) \;\middle|\; u' \in L^q(I; X_1) \,\right\}.$$

Then the following holds:

(i) For $p < \infty$ *the embedding* $W_{p,q}(I; X_0, X_1) \hookrightarrow L^p(I; X)$ *is compact and*

(ii) for $q > 1$ *the embedding* $W_{\infty,q}(I; X_0, X_1) \hookrightarrow C(I; X)$ *is compact.*

Proof. See [Aub63; Sim86]. □

We are mostly interested in the special case that the spaces in the lemma are equal to our Gelfand-Triple $V \subset H \subset V^*$. By induction we get the following result for functions of higher regularity in time.

Corollary 4.15. *Given* p, q *with* $1 \leqslant p \leqslant \infty$, $1 \leqslant q \leqslant p$ *and* $q < \infty$, *both of the embeddings*

(i) $W^{k,p}(I; V) \cap W^{k+1,p}(I; H) \hookrightarrow W^{k,q}(I; H)$ *and*

(ii) $W^{k,\infty}(I; V) \cap W^{k+1,\infty}(I; H) \hookrightarrow C^k(I; H)$

are compact.

We apply this to the derivatives of S and obtain the following.

Lemma 4.16. *Let* $k \in \mathbb{N} \cup \{0\}$, $f \in \mathcal{F}^{(k)}$ *and* $u_0 = u_1 = 0$ *if* $k \geqslant 1$. *Further let* $p \in D(S) \cap X^{(k)}$ *and* $k_1 := \max\{1, k\}$. *Then the following holds.*

(i) For $S\colon D(S) \cap X^{(k)} \to Z$ *with* $Z = W^{j,p}(I; H)$ *or* $Z = C^j(I; H)$ *with* $0 \leqslant j \leqslant k$ *and* $1 \leqslant p < \infty$, *its derivatives*

$$\begin{aligned} \partial_Q S(p) &\in \mathcal{L}(W^{k_1,\infty}(I; \mathcal{L}(V, H)), Z) \quad \textit{and} \\ \partial_B S(p) &\in \mathcal{L}(W^{k_1,\infty}(I; \mathcal{L}(H)), Z) \end{aligned}$$

are compact operators.

(ii) If $k \geqslant 2$ *and* $S\colon X^{(k)} \to Z$ *with* $Z = W^{j,p}(I; H)$ *or* $Z = C^j(I; H)$ *with* $0 \leqslant j \leqslant k-1$ *and* $1 \leqslant p < \infty$, *then the operators*

$$\begin{aligned} \partial_A S(p) &\in \mathcal{L}(W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*)), Z) \quad \textit{and} \\ \partial_C S(p) &\in \mathcal{L}(W^{k+1,\infty}(I; \mathcal{L}^{sa}(H)), Z) \end{aligned}$$

are compact as well.

Proof. (i) Follows from the compactness of $Y^{(k)} \hookrightarrow Z$.

(ii) We observe that the space $Y^{(k-2)}$ is continuously embedded in Z, i.e. $S\colon X^{(k)} \to Z$ is in fact Fréchet-differentiable with respect to A and C. Moreover, $\partial_A S(p)$ and $\partial_C S(p)$ map into $Y^{(k-1)}$, which has a compact embedding into Z. □

From the compactness of the derivatives we know that the linearized problems arising from S would be locally ill-posed at every point. However, they might still be locally well-posed after restricting the problem to $(\ker \partial_x S(p))^\perp$. In other words, the corresponding generalized inverses might still be continuous, yielding well-posed problems. For a compact linear operator this can only hold if its image is finite-dimensional. We show that this is not the case here.

Lemma 4.17. *Assume everything as in Lemma 4.16 and additionally suppose that* $f \neq 0$. *In this setting, the range of the following operators is infinite-dimensional for every* $p \in D(S) \cap X^{(k)}$*:*

(i) $\partial_Q S(p) \in \mathcal{L}(W^{k_1,\infty}(I;\mathcal{L}(V,H)), Y^{(k)})$,

(ii) $\partial_A S(p) \in \mathcal{L}(W^{k+1,\infty}(I;\mathcal{L}^{sa}(V,V^*)), Y^{(k-1)})$ *if* $k \geqslant 2$,

(iii) $\partial_B S(p) \in \mathcal{L}(W^{k,\infty}(I;\mathcal{L}(H)), Y^{(k)})$ *if* $k \geqslant 1$ *and*

(iv) $\partial_C S(p) \in \mathcal{L}(W^{k+1,\infty}(I;\mathcal{L}^{sa}(H)), Y^{(k-1)})$ *if* $k \geqslant 2$.

Proof. In contrast to all previous proofs, this one is more naturally divided into the cases A & Q and B & C.

(i) & (ii): Assume that one of the operators had a finite-dimensional range, i.e. that $u_h = \partial_x S(p)[h]$ (with $x = A$ or $x = Q$) could be represented as a finite sum independent of $h \in W^{k,\infty}(I;\mathcal{L}(V,H))$ and $h \in W^{k+1,\infty}(I;\mathcal{L}^{sa}(V,V^*))$, respectively. Due to the linearity of the equation solved by u_h its left- and therefore also its right-hand side $-h[u]$ could then be written as a finite sum as well. To be more precise, we would have

$$-h[u] = \sum_{j=1}^{M} \alpha_j(h)\left[(\mathcal{C}\psi_j')' + \mathcal{B}\psi_j' + (\mathcal{Q}+\mathcal{A})\psi_j\right]. \tag{4.21}$$

We proceed to show that this cannot be the case. Since $f \neq 0$ we also have $u = S(p) \neq 0$ and (even for $k = 0$) know that $u \in C(I;H)$. Thus there exist $t_0 \in (0,T)$ and $\varepsilon > 0$ such that $t_0 + (-\varepsilon,\varepsilon) \subset (0,T)$ and $u(t_0 + s) \neq 0$ for all $s \in (-\varepsilon,\varepsilon)$. Given any sequence of pointwise disjoint balls $B(t_i,\varepsilon_i) \subset t_0 + (-\varepsilon,\varepsilon)$ and functions $(\beta_i)_{i\in\mathbb{N}} \subset C^\infty(\mathbb{R})$ with $\emptyset \neq \operatorname{spt}\beta_i \subset B(t_i,\varepsilon_i)$, we define $h_i(t) \coloneqq \beta_i(t)\mathrm{Id}_H$. This way we get $h_i \in C^\infty(I;\mathcal{L}(H)) \subset C^\infty(I;\mathcal{L}^{sa}(V,V^*))$. The supports of $-h_i[u]$ are nonempty and pairwise disjoint. Hence, the set $\{-h_i[u]\}_{i\in\mathbb{N}}$ is infinite and linearly independent, which contradicts (4.21).

(iii) & (iv): Here h is applied to u'. To have $u' \in C(I;H)$ in order to make a similar argument as in the previous case, we have to require regularity with $k \geqslant 1$. Due to $u_0 = 0$ and $u \neq 0$, we conclude $u' \neq 0$ and can proceed as in the first part of the proof and obtain $\dim(\mathcal{R}(\partial_B S(p))) = \infty$. When looking at $\partial_C S$, we additionally have to choose β_i in such a way that $(\beta_i u')' = \beta_i' u' + \beta_i u'' \neq 0$. This poses no difficulties for $k \geqslant 2$, because in this case u'' is continuous as well. $\square$

5 APPLICATION TO THE ACOUSTIC WAVE EQUATION

As the first and main application of the abstract theory developed in Chapter 4, we consider the acoustic wave equation. For a *wave speed* c and a *mass density* ρ that do not depend on time, a variant of this equation reads

$$\frac{1}{c(x)^2\rho(x)}u''(t,x) - \operatorname{div}\frac{\nabla u(t,x)}{\rho(x)} = f(t,x). \tag{5.1}$$

Together with suitable initial- and boundary conditions it serves as a simple model for the propagation of acoustic waves in fluids [CK13; Ika00; Jen+11; Kir11] and also seismic waves [KR14a; Sym09]. In these cases the unknown u in the equation above is the *acoustic pressure*. The differential equation (5.1) is not entirely new to us; we have already encountered a simplified version of it in Examples 2.4, 2.11 and 3.11.

From a modeling viewpoint it is not clear at all, whether it is advisable to also employ the partial differential equation (5.1) for dynamic wave speed and density. The position of static parameters c, ρ in relation to the time derivatives can be changed without changing the equation, but for time-dependent parameters this is obviously not the case. Hence, we are going to start this chapter by taking a look at one possible derivation of the acoustic wave equation. This will include the comparison of the different possible second-order terms for the time variable. Since we are not bound to a specific application, and mainly want to showcase the power of our general framework as detailed as possible, we will add two additional parameters to the equation. The result of Section 5.1 will be a boundary value problem for the equation

$$\frac{1}{\rho}\left(\frac{1}{c^2}u'\right)' + \nu u' - \operatorname{div}\frac{\nabla u}{\rho} + qu = f, \tag{5.2}$$

which gives rise to multiple inverse problems, one for each of the four time- and space-dependent coefficients on the left-hand side of this equation. The analysis of these inverse problems in the context of Chapter 4's abstract framework will then be the primary focus of the remainder of this chapter. To this end we first restate the acoustic wave equation as an evolution equation in Section 5.2, and also set up a *value operator* (in the spirit of Figure 4.1 on page 44) that maps the unknown parameters onto the abstract operators A, B, C and Q. This operator can then be composed with the forward operator from the general theory. This also allows showing Fréchet-differentiability of the parameter-to-solution map in the subsequent Section 5.3, where we further give a characterization of the adjoint of this derivative. We conclude this chapter by proving the ill-posedness of the corresponding inverse problems in Section 5.4.

These theoretical consideration will be put to practical use in Chapters 7 and 8, where we discretize (5.2) and tackle the numerical reconstruction of each of the four parameters.

5.1 PURSUIT OF A SUITABLE EQUATION

To obtain the acoustic wave equation in a three-dimensional setting, we proceed like JENSEN et al. [Jen+11]. We assume that the acoustic waves have a small amplitude, which justifies the linearization of the acoustic pressure p, the particle velocity field v and the mass density ρ at the ambient state $p_0 \in \mathbb{R}$, $v_0 = 0 \in \mathbb{R}^3$ and $\rho_0 \in \mathbb{R}$. The small deviations from these background values are denoted by p_1, v_1 and ρ_1. For these, we have the linearized continuity equation

$$\rho_1' = -\operatorname{div}(\rho_0 v_1), \tag{5.3}$$

the linearized Euler equation

$$v_1' = -\frac{\nabla p_1}{\rho_0} \tag{5.4}$$

and the linearized equation of state

$$p_1' = c^2(\rho_1' + v_1 \cdot \nabla \rho_0). \tag{5.5}$$

At least for now, the quantity c in the state equation is some constant that depends on the material. We differentiate the equation of continuity (5.3) with respect to time, and assume that $\rho_0' \operatorname{div} v_1$ is sufficiently small to justify

$$\begin{aligned}\rho_1'' &= -(\operatorname{div}(\rho_0 v_1))' = -(\rho_0 \operatorname{div} v_1)' - (\nabla \rho_0 \cdot v_1)' \\ &\approx -\rho_0 \operatorname{div} v_1' - (\nabla \rho_0 \cdot v_1)'.\end{aligned}$$

From the Euler equation (5.4) we obtain $\operatorname{div} v_1' = -\operatorname{div} \frac{\nabla p_1}{\rho_0}$, which we insert into the previous equation to see

$$\rho_1'' = \rho_0 \operatorname{div} \frac{\nabla p_1}{\rho_0} - (\nabla \rho_0 \cdot v_1)'. \tag{5.6}$$

The equation of state (5.5) can be solved for ρ_1'. Together with (5.6) this yields

$$\left(\frac{1}{c^2} p_1'\right)' - (\nabla \rho_0 \cdot v_1)' = \rho_0 \operatorname{div} \frac{\nabla p_1}{\rho_0} - (\nabla \rho_0 \cdot v_1)'.$$

By adding $(v_1 \cdot \nabla \rho_0)'$ to both sides we end with the wave equation

$$\frac{1}{\rho_0}\left(\frac{1}{c^2} p_1'\right)' - \operatorname{div} \frac{\nabla p_1}{\rho_0} = 0 \tag{5.7}$$

for the perturbation p_1 of the acoustic pressure p.

We were correct in assuming that (5.1) is not appropriate for time-dependent wave speed c because c^{-2} has to be *between* the two time derivatives that act on p_1. In contrast to this, the position of ρ^{-1} remains unchanged, i.e. it belongs outside of both time derivatives.

5.1.1 *The wave speed*

The constant c in the wave equation is referred to as the *wave speed*, because at least for time-independent c it determines how fast the solution u of (5.7) propagates. We would like to discuss the *qualitative* effect a time-dependent c has on the wave, and how it has to enter the wave equation in order to describe the speed of the waves.

The fact that a constant $c > 0$ is in fact the wave speed can easily be verified for the one-dimensional Cauchy-Problem

$$\begin{aligned} \frac{1}{c^2}u''(t,x) - \partial_x^2 u(t,x) &= 0, \qquad (t,x) \in (0,\infty) \times \mathbb{R}, \\ u(0,x) = u_0(x) \ \text{ and } \ u'(0,x) &= 0, \qquad x \in \mathbb{R}, \end{aligned}$$

with given initial values $u_0 \in C^2(\mathbb{R})$. It is well-known that a classical solution $u \in C^2((0,\infty) \times \mathbb{R})$ of this problem exists and that it is for all $t > 0$ and $x \in \mathbb{R}$ given by the *d'Alembert formula*

$$u(t,x) = \frac{1}{2}u_0(x - ct) + \frac{1}{2}u_0(x + ct). \tag{5.8}$$

This solution formula shows that the initial values u_0 are separated into a left- and a right traveling wave, each moving with speed c. There are analogous formulas for the case of more than one space dimension. In two dimensions it is called *Kirchhoff formula*, and the variant for three dimensions is known as the *Poisson formula*, see for example [Eva10]. The most striking implication of these closed formulas for the solution u is that they reveal a fundamental difference in wave propagation in different space dimensions d: For odd d, the value of $u(t,x)$ only depends on the values of u_0 on the sphere[1] $\partial B(x,ct)$. In contrast, for even d the value $u(t,x)$ takes the initial values in the whole ball $B(x,ct)$ into account.

If the initial values in our one-dimensional example satisfy $\operatorname{spt} u_0 = (-\infty, 0]$, then $\operatorname{spt} u(t) = (-\infty, ct]$ and therefore the volume of the support of $u(t)$ grows at the rate c. If we define the wave speed based on this observation, then the extension to time-dependent wave speeds happens quite naturally by defining

$$c(t) \coloneqq \frac{d}{dt}\operatorname{vol}(\operatorname{spt} u(t)). \tag{5.9}$$

In case that we want energy transporting solutions like (5.8), we should require that solutions have the form $u(t,x) = \frac{1}{2}u_0(x - g(t)) + \frac{1}{2}u_0(x + g(t))$. Then the wave speed is $c(t) = g'(t)$ and the initial condition $u'(0) = 0$ is automatically satisfied. For the other initial condition $u(0) = u_0$ to hold, we have to assume $g(0) = 0$, hence $g(t) = \int_0^t c(s)\,ds$. This ansatz for the solution u solves the partial differential equation

$$\frac{1}{c(t)}\left(\frac{1}{c(t)}u'(t,x)\right)' - \partial_x^2 u(t,x) = 0. \tag{5.10}$$

1 This is Huygen's principle: Every point in the *wave front* is the origin of a new spherical wave.

Of course we do not claim that solutions to (5.10) are the only ones traveling at speed c because for its derivation we have also assumed conservation of energy in u, and it is not unreasonable to assume that a change in c changes the energy in the system.

Alongside the "mixed" approach of (5.10), there are two other options of how to include the parameter c in the equation that we wish to discuss. First, using the leading term $(u'/c^2)'$, like the wave equation (5.7) we just derived. Second, leaving (5.1) as-is, which means having c outside of both time derivatives. It is difficult to reason about these other possibilities of including the wave speed in the equation because we do not have closed solution formulas for them. However, from the theory of ordinary differential equations we know that coefficients in front of the first-order time derivative of u only contribute to a damping or amplification of the wave, depending on the sign of the coefficient (see [EGK17] for more details). This suggests that the speed of the wave propagation is solely determined by the term u'' and factors in front of it. From the equalities

$$\begin{aligned}\frac{u''(t,x)}{c(t)^2} &= \frac{1}{c(t)}\frac{d}{dt}\left(\frac{u'(t,x)}{c(t)}\right) + \frac{c'(t)}{c(t)^3}u'(t,x)\\ \frac{d}{dt}\left(\frac{u'(t,x)}{c(t)^2}\right) &= \frac{1}{c(t)}\frac{d}{dt}\left(\frac{u'(t,x)}{c(t)}\right) - \frac{c'(t)}{c(t)^3}u'(t,x)\end{aligned}$$

we conclude that the propagation rate of the wave should be the same in each case. Moreover, it follows that we can view the solutions u_{in} and u_{out} of

$$\frac{d}{dt}\left(\frac{u_{\mathrm{in}}'}{c^2}\right) - \partial_x^2 u_{\mathrm{in}} = \frac{1}{c}\frac{d}{dt}\left(\frac{u_{\mathrm{in}}'}{c}\right) - \frac{c'}{c^3}u_{\mathrm{in}}' - \partial_x^2 u_{\mathrm{in}} = 0, \tag{5.11a}$$

$$\frac{u_{\mathrm{out}}''}{c^2} - \partial_x^2 u_{\mathrm{out}} = \frac{1}{c}\frac{d}{dt}\left(\frac{u_{\mathrm{out}}'}{c}\right) + \frac{c'}{c^3}u_{\mathrm{out}}' - \partial_x^2 u_{\mathrm{out}} = 0, \tag{5.11b}$$

in $(0,\infty)\times\mathbb{R}$ that also satisfy the initial conditions

$$u_{\mathrm{in}}(0) = u_{\mathrm{out}}(0) = u_0, \quad u_{\mathrm{in}}'(0) = u_{\mathrm{out}}'(0) = 0$$

as (locally) damped or amplified versions of the energy conserving solution u of (5.10). An increase in $c(t)$ would at least locally result in a damping of $u_{\mathrm{out}}(t)$ and an amplification of $u_{\mathrm{in}}(t)$. According to (5.7), the latter is the more "physical" behavior.

WAVE SPEED IN SIMULATIONS We would like to reinforce this argument by providing some numerical examples, in which we measure the time it takes *approximations* of the functions u, u_{in} and u_{out} to propagate a set distance. To minimize oscillations due to numerical dispersion we use smooth initial values, given by

$$u_0(x) := \begin{cases}\exp\left(1 - \frac{1}{1-x^2}\right) & \text{if } |x| < 1,\\ 0 & \text{otherwise.}\end{cases} \tag{5.12}$$

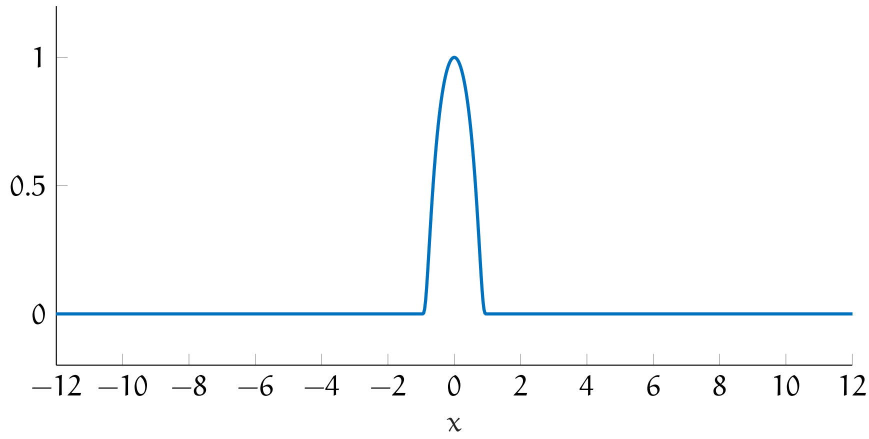


Figure 5.1: Initial values u_0 as defined in (5.12)

A plot of this function is depicted in Figure 5.1.

Because we are faced with only one space dimension in this example, it is convenient to employ finite differences to discretize the PDEs. We want to see how long it takes for the wave to leave the interval $(-11, 11)$, therefore it is sufficient if we restrict the discretization to $\Omega = [-12, 12]$. Furthermore, we can pose homogeneous Dirichlet boundary conditions at $x = -12$ and $x = 12$ without changing the solution in the time interval of interest. For the spatial discretization we employ an equidistant mesh of Ω with 1024 nodes, which means the distance between nodes is $\Delta x \approx 0.03$. The time discretization is performed using Crank-Nicolson, with the step size $\Delta t := \Delta x/16 \approx 1.8 \cdot 10^{-3}$. The Crank-Nicolson method will be introduced in Chapter 7 in more detail. In each step we have to solve a 1024×1024 linear system. Because of this small size we can employ a direct solver for this task. The parameter c that we use in this experiment is

$$c(t) = 2 + \cos t, \qquad t \geqslant 0. \tag{5.13}$$

Hence, the support of the exact solution u of (5.10) should leave $(-11, 11)$ at t^*, which is the solution of

$$10 = \int_0^{t^*} c(s)\,ds = 2t^* + \sin(t^*).$$

From the positivity of c we conclude that t^* is unique and we also immediately see that $t^* \in (4.5, 5.5)$. An application of Newton's method yields the approximation $t^* \approx 5.39$.

We stop the time integration as soon as the wave field evaluated at $x = 11$ is greater than one percent of its current maximal value in $(-12, 12)$. For the approximations of u_{in} and u this is the case at $t \approx 5.39$, and the absolute difference of this value to (the approximation of) t^* is $7.5 \cdot 10^{-4}$. Note that this error is smaller than the time step size Δt. The field u_{out} satisfies the stopping criterion exactly one time step later. We conclude that the arrival times are independent of the choice of the model and are equal to the expected value t^*.

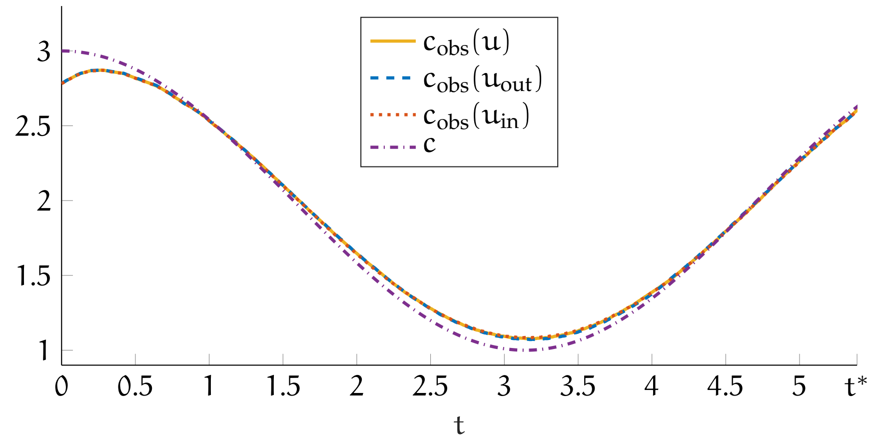


Figure 5.2: Observed wave speeds of u_{in}, u_{out} and u, together with the parameter c as given in (5.13)

In theory, we should also be able to compute the propagation speed of the waves (in the sense of (5.9)) throughout the solution process by observing the rate at which the support of the solution grows. Since the waves can propagate in two directions, we have to supplement (5.9) with the factor $1/2$, i.e.

$$[c_{obs}(v)](t) \coloneqq \frac{1}{2}\frac{d}{dt}\operatorname{vol}(\operatorname{spt} v(t)).$$

However, the discrete nature of $\operatorname{vol}(\operatorname{spt} v(t))$ for a numerically approximated v makes the calculation of c_{obs} difficult: The support's size can only increase by a multiple of Δx and in fact does so only every few time steps, yielding a heavily stair-cased function that we subsequently have to differentiate. We interpret this as an ill-posed inverse problem with the forward operator $A \in \mathcal{L}(L^2([0,t^*]))$, $(Af)(t) = \int_0^t f(s)\,ds$, which we approximate using an explicit euler scheme and then invert by employing Tikhonov-Regularization.[2] For the three functions u_{in}, u_{out} and u, this results in observed speeds $c_{obs}(t)$ as depicted in Figure 5.2, which shows that the propagation speed of all functions are identical. Most importantly, these speeds are reasonably close to the time-dependent parameter c to warrant naming it the *wave speed*.

Nevertheless, the models exhibit different solution behavior, which we discuss using the visualizations of their solutions at t^*, which can be found in Figure 5.3. It is clearly visible that u has the expected behavior of splitting the initial values into two wave fronts that travel through the domain. For u_{in} and u_{out} this is not the case; these solutions do not return to zero after the leading wave front has passed. We note that the coefficient c'/c^3 is negative in $(0,\pi)$ and positive in $(\pi, 2\pi)$. Since $t^* \in (\pi, 2\pi)$, the PDEs (5.11) suggest that u_{out} should have experienced

2 By doing so we incidentally solve a simplified version of one of our main inverse problems: The determination of a purely time-dependent wave speed from the solution u, at least in a one-dimensional setting. It is good to know that even this simplified task requires some thought, and in particular, regularization.

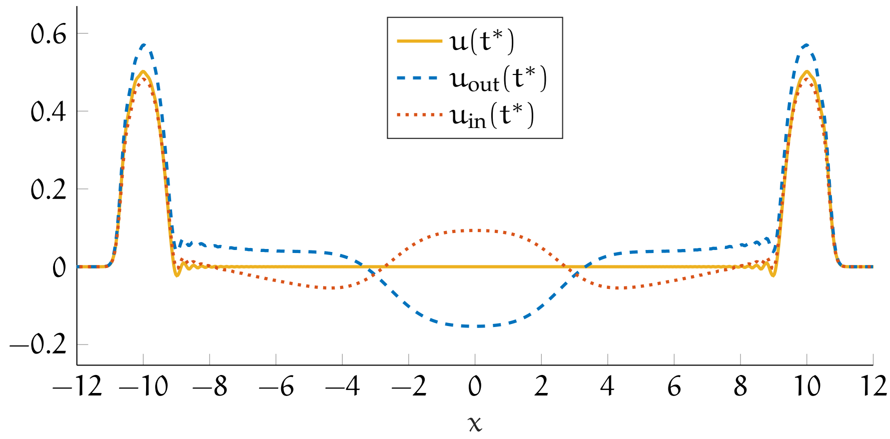


Figure 5.3: Approximations of u_{in}, u_{out} and u as they leave $(-11, 11)$

more amplification than damping in the time window $(0, t^*)$ and that the opposite should be the case for u_{in}. This indicates why the maximal values of $u_{out}(t^*)$ and $u_{in}(t^*)$ are greater than $1/2$ and less than $1/2$, respectively.

The preceding discussion shows that the parameter c in the wave equation $\rho^{-1}(c^{-2}u')' - \operatorname{div}(\rho^{-1}\nabla u) = f$, which we derived in the context of acoustic waves, is only one of the possible ways to include the wave speed in the model. However, due to the position of ρ^{-1} we will also see how to deal with coefficients outside of the time derivatives. Thus, we can safely decide to use this equation without loss of generality, since the analysis in the remainder of this chapter can easily be modified to account for a different position of c by treating it in a similar way as ρ.

ADDITIONAL PARAMETERS Moreover, since our main goal is not a specific application, but the showcasing of the identification of time-dependent parameters, we choose to augment (5.7) by two additional parameters: First, the coefficient q in the zero-order term qu. The identification of such a parameter was the sole task in [GL17]. A coefficient like this arises for example when dealing with the linearization of a previously nonlinear wave equation. Second, the parameter ν in $\nu u'$. We already know that such a term induces a damping- or amplification of the wave field u. We note that our theory would also allow to add a transportation term $e \cdot \nabla u$ with a vector-valued parameter e to the equation, but we will not do so in order to not unnecessarily complicate the presentation. All in all, we obtain the wave equation

$$\frac{1}{\rho}\left(\frac{1}{c^2}u'\right)' + \nu u' - \operatorname{div}\frac{\nabla u}{\rho} + qu = f.$$

With it come the four inverse problems of determining the time- and space-dependent coefficients c, ν, ρ or q from knowledge of u. To make the application of the abstract framework easier, we equip u with homogeneous initial conditions from the start, despite the fact that for results concerning q and ν (that need no higher regularity) this is not necessary. Furthermore, we assume that the excitation of waves only happens through the right-hand side f by imposing homogeneous Dirichlet boundary conditions on u. In the context of acoustic waves this means that the domain is surrounded by immovable walls.

5.2 CONSTRUCTION OF THE FORWARD OPERATOR

Let $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^d$ a bounded domain and $I = [0, T]$ for some $T > 0$. Motivated by the preceding section, we concentrate on the initial boundary value problem

$$\frac{1}{\rho}\left(\frac{1}{c^2}u'\right)' + \nu u' - \operatorname{div}\frac{\nabla u}{\rho} + qu = f \qquad \text{in } (0,T)\times\Omega, \tag{5.14a}$$

$$u(0) = \left(c^{-2}\rho^{-1}u'\right)(0) = 0 \qquad \text{in } \Omega, \tag{5.14b}$$

$$u = 0 \qquad \text{on } (0,T)\times\partial\Omega. \tag{5.14c}$$

The wave field u, the right-hand side f and the four unknown coefficients c, ν, ρ and q are assumed to be real functions on $I \times \Omega$. Note that even though u is now a function of two sets of variables t and x, we continue to write $u(t)$ when we actually mean $u(t, \cdot)$. The formulation of the second initial condition in (5.14b) already keeps in mind that the weak solution to (5.14) might not yield a well-defined $u'(0)$. However, if the solution u and the coefficients c, ρ are regular enough, then the solution will of course equivalently satisfy $u'(0) = 0$.

Due to the presence of ρ^{-1} outside of the time derivatives in the leading term, this problem does not immediately yield an evolution equation of the form $(\mathcal{C}u')' + \mathcal{B}u' + \mathcal{A}u + \mathcal{Q}u = f$. Instead, it is more naturally stated as an equation in which the leading term is replaced by $\mathcal{D}(\mathcal{C}u')'$. We know from Corollary 3.12 that such an equation can be solved when assuming that the coefficients are smooth because then it is equivalent to the evolution equation

$$(\mathcal{D}\mathcal{C}u')' + (\mathcal{B} - \mathcal{D}'\mathcal{C})u' + (\mathcal{A} + \mathcal{Q})u = f.$$

However, we do not want to start out with these extra smoothness assumptions, because it needlessly causes the results for the other parameters to get worse. Instead, we base our analysis on the re-stated equation, bearing in mind that it is only equivalent to the original equation (5.14a) if $c^{-2}u' \in H^1(I; L^2(\Omega))$ (regularity with $k = 1$) or if ρ does not depend on time. In the general operator framework this reformulation was carried out by using our product rule Lemma 2.3. In the context of a specific partial differential equation, this boils down to the classical product rule, which yields

$$\frac{1}{\rho}\left(\frac{1}{c^2}u'\right)' = \left(\frac{1}{\rho c^2}u'\right)' - \frac{\rho'}{\rho^2 c^2}u',$$

at least as long as $c^{-2}u'$ is weakly differentiable in time. Hence, we simply replace (5.14) by

$$\left(\frac{1}{\rho c^2}u'\right)' + \left(\frac{\rho'}{\rho^2 c^2} + \nu\right)u' - \operatorname{div}\frac{\nabla u}{\rho} + qu = f \qquad \text{in } (0,T)\times\Omega, \tag{5.15a}$$

$$u(0) = \left(c^{-2}\rho^{-1}u'\right)(0) = 0 \qquad \text{in } \Omega, \tag{5.15b}$$

$$u = 0 \qquad \text{on } (0,T)\times\partial\Omega. \tag{5.15c}$$

Due to the boundary conditions, the suitable function space for $u(t)$ is $V := H^1_0(\Omega)$. By setting $H := L^2(\Omega)$ we obtain the Gelfand triple

$$H^1_0(\Omega) \subset L^2(\Omega) \subset H^{-1}(\Omega)$$

that has all of the properties we asserted in Definition 2.1. In this context this implies that every function belonging to $L^2(\Omega)$ or $H^1_0(\Omega)$ becomes an element of $H^{-1}(\Omega)$ through the inner product of $L^2(\Omega)$. To ease notation, we will sometimes omit the "(Ω)" part for Lebesgue- or Sobolev spaces connected to the domain Ω whenever the expressions tend to become unwieldy, for example when they appear in Bochner spaces or in subscripts.

The weak formulation of (5.15) is obtained by integrating over the domain and then formally integrating by parts. It reads

$$\begin{aligned}\frac{d}{dt}\left(C_{c,\rho}(t)u'(t),\varphi\right) + \left(B_{c,\nu,\rho}(t)u'(t),\varphi\right) \\ + \langle (A_\rho(t) + Q_q(t))u(t),\varphi\rangle = \langle f(t),\varphi\rangle\end{aligned} \tag{5.16}$$

for all $\varphi \in H^1_0(\Omega)$, which in turn should be fulfilled for almost all $t \in I$. We seek a solution $u \in L^2(I; H^1_0(\Omega)) \cap H^1(I; L^2(\Omega))$ that additionally adheres to the initial conditions $u(0) = 0$ (as an equality in $L^2(\Omega)$) and $(\mathcal{C}_{c,\rho}u')(0) = 0$ (holding in $H^{-1}(\Omega)$). As in the abstract framework, we denote with $(\cdot,\cdot)$ the inner product of $H = L^2(\Omega)$ and with $\langle\cdot,\cdot\rangle$ the dual pairing of $V^* = H^{-1}(\Omega)$ and $H^1_0(\Omega)$. The operators that appear in the weak formulation are defined for $t \in I$, $\varphi, \psi \in H^1_0(\Omega)$ and $v \in L^2(\Omega)$ by

$$\langle A_\rho(t)\psi,\varphi\rangle := \int_\Omega \frac{\nabla\psi(x)\cdot\nabla\varphi(x)}{\rho(t,x)}\,dx, \qquad C_{c,\rho}(t)v := \frac{1}{\rho(t)c(t)^2}v,$$

$$B_{c,\nu,\rho}(t)v := \left(\frac{\rho'(t)}{\rho(t)^2 c(t)^2} + \nu(t)\right)v, \qquad Q_q(t)v := q(t)v.$$

With "$\cdot$" we denote the usual inner product of $\mathbb{R}^d$. We see that except for $A(t)$, all of the operators are multiplication operators; for example $Q_q(t)$ multiplies its argument with the space-dependent function $q(t) = q(t,\cdot)\colon \Omega \to \mathbb{R}$. By making use of distributional derivatives, we can also write $A_\rho(t)\psi = -\operatorname{div}\left(\rho(t)^{-1}\nabla\psi\right) \in H^{-1}(\Omega)$, which avoids the appearance of the test function φ.

By utilizing Lemma 2.2, we see that the weak formulation (5.16), combined with the homogeneous initial conditions (5.15b), is equivalent to the evolution problem

$$(\mathcal{C}_{c,\rho}u')' + \mathcal{B}_{c,\nu,\rho}u' + (\mathcal{A}_\rho + \mathcal{Q}_q)u = f \quad \text{in } L^2(I;H^{-1}(\Omega)), \tag{5.17a}$$

$$u(0) = 0 \;\text{ and }\; (\mathcal{C}_{c,\rho}u')(0) = 0, \tag{5.17b}$$

which is exactly the problem that we thoroughly analyzed in Chapter 4.

The forward operator to our inverse problems reads

$$F(c,\nu,\rho,q) := u,$$

where u solves (5.17). In Chapter 4 we assumed that such a forward operator could be decomposed into the operator S, that maps the operators A, B, C and Q onto u, preceded by a value operator P that maps the unknown parameters onto the four operators (cf. Figure 4.1). We immediately see $F = S \circ P$ holds, with

$$P(c,\nu,\rho,q) := (A_\rho, B_{c,\nu,\rho}, C_{c,\rho}, Q_q). \tag{5.18}$$

We would like to note that this is the most general way to set up the value operator P, because it does not include any a-priori information about the parameters. If more information is available, e.g. that the inhomogeneities in Ω simply move through the domain at a constant speed, then this could also be modeled through P. In this hypothetical case, one might use four time-independent functions c, ν, ρ and q as the unknowns, together with a single direction vector $v \in \mathbb{R}^d$ that controls the movement.

In order to formally set up F, we need to find a domain of definition for P that guarantees that it maps to $D(S)$, or even $D(S) \cap X^{(k)}$ whenever we require a more regular solution u, since we have $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$. The spaces $X^{(k)}$ and $Y^{(k)}$ were introduced in equations (4.2) and (4.3) on page 47. Note that in the setting at hand, the operator $Q(t)$ belongs to $\mathcal{L}(H)$. In Chapter 4 we allowed it to be more general by assuming $Q(t) \in \mathcal{L}(V,H)$. In return, we had to assume differentiability of Q even in the space $X = X^{(0)}$. This assumption was only needed to ensure the uniqueness of the solution through Theorem 2.10, therefore we can drop this requirement without impairing any of the results we have for S. All in all, this means that we will make use of the spaces

$$\begin{aligned} X^{(k)} &:= W^{k+1,\infty}\big(I;\mathcal{L}^{sa}(H_0^1,H^{-1})\big) \times W^{k_1,\infty}\big(I;\mathcal{L}(L^2)\big) \\ &\quad\; \times W^{k+1,\infty}\big(I;\mathcal{L}^{sa}(L^2)\big) \times W^{k,\infty}\big(I;\mathcal{L}(L^2)\big) \\ \text{and }\; Y^{(k)} &:= W^{k,\infty}(I;H_0^1) \cap W^{k+1,\infty}(I;L^2) \end{aligned}$$

for $k \in \mathbb{N}_0$ and $k_1 := \max\{1,k\}$. The open domain of definition of S is as before given as

$$\begin{aligned} D(S) := \Big\{ (A,B,C,Q) \in X^{(0)} \;\Big|\; & A(t) \in \mathcal{L}^{sa}_{A_0+\varepsilon}\big(H_0^1(\Omega),H^{-1}(\Omega)\big) \text{ and} \\ & C(t) \in \mathcal{L}^{sa}_{C_0+\varepsilon}\big(L^2(\Omega)\big) \text{ for almost all } t \in I \text{ for some } \varepsilon > 0 \Big\} \end{aligned}$$

and ensures that A and C are not only self-adjoint, but also coercive.

The only thing that is missing for a well-defined $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$ is the assumption that the right-hand side f is either an element of $H^k\big(I; L^2(\Omega)\big)$ or $H^{k+1}\big(I; H^{-1}(\Omega)\big)$ and it has a root of sufficient high order at $t = 0$. In equation (4.4) we introduced the set $\mathcal{F}^{(k)}$ that incorporates these properties.

We proceed by setting up the value operator P from (5.18) in a way that its image is a subset of $D(S) \cap X^{(k)}$. Clearly, differentiability of the parameters (c, ν, ρ, q) directly translates to differentiability of $P(c, \nu, \rho, q)$ with respect to time. To ensure that $A_\rho(t)$ and $C_{c,\rho}(t)$ are well-defined we assume that $\rho(t, x) \geqslant \rho_0 > 0$ and $c(t, x) \geqslant c_0 > 0$ hold for almost all $(t, x) \in I \times \Omega$. Note that both operators are self-adjoint. Regarding their coercivity, we observe that if $\rho(t, x) \leqslant \rho_1 < \infty$ and $c(t, x) \leqslant c_1 < \infty$ for almost all $(t, x) \in I \times \Omega$, then

$$(C_{c,\rho}(t)v, v) = \int_\Omega \frac{v(x)^2}{\rho(t, x)c(t, x)^2}\,dx \geqslant \rho_1^{-1} c_1^{-2} \|v\|^2_{L^2(\Omega)}$$

holds for all $v \in L^2(\Omega)$. Since Ω is bounded, it provides a Poincaré inequality of the form $\|\psi\|_{H^1_0(\Omega)} \leqslant C_p \|\nabla\psi\|_{L^2(\Omega;\mathbb{R}^d)}$ with a constant $C_p > 0$. Hence,

$$\langle A_\rho(t)\psi, \psi\rangle = \int_\Omega \frac{|\nabla\psi(x)|^2}{\rho(t, x)}\,dx \geqslant \rho_1^{-1} C_p^{-2} \|\psi\|^2_{H^1_0(\Omega)}$$

is valid for all $\psi \in H^1_0(\Omega)$. Let the constants ρ_0, ρ_1, c_0 and c_1 be fixed throughout the chapter. The previous considerations motivate the definitions

$$\begin{aligned} \mathcal{W}^{(k)} := {} & W^{k+1,\infty}(I; L^\infty(\Omega)) \times W^{k_1,\infty}(I; L^\infty(\Omega)) \\ & \times W^{k+1,\infty}(I; L^\infty(\Omega)) \times W^{k,\infty}(I; L^\infty(\Omega)), \end{aligned} \tag{5.19}$$

$$\begin{aligned} D(P) := \Big\{ (c, \nu, \rho, q) \in \mathcal{W}^{(0)} \;\Big|\; & \rho_0 + \varepsilon \leqslant \rho \leqslant \rho_1 - \varepsilon \text{ and} \\ & c_0 + \varepsilon \leqslant c \leqslant c_1 - \varepsilon \;\text{ a.e. in } I \times \Omega \text{ for a } \varepsilon > 0 \Big\}, \end{aligned} \tag{5.20}$$

because in this way we obtain a well-defined

$$\begin{aligned} P\colon D(P) \cap \mathcal{W}^{(k)} &\to D(S) \cap X^{(k)}, \\ (c, \nu, \rho, q) &\mapsto (A_\rho, B_{c,\nu,\rho}, C_{c,\rho}, Q_q). \end{aligned}$$

For this we set the constants in $D(S)$ to be $C_0 := \rho_1^{-1} c_1^{-2}$ and $A_0 := \rho_1^{-1} C_p^{-2}$. Moreover, by the same reasoning as in Lemma 4.2, $D(P) \cap \mathcal{W}^{(k)}$ forms an open subset of $\mathcal{W}^{(k)}$ for all $k \in \mathbb{N}_0$.

The forward operator $F = S \circ P$ can then be viewed as

$$\begin{aligned} F\colon D(F) \cap \mathcal{W}^{(k)} &\to Y^{(k)}, \\ (c, \nu, \rho, q) &\mapsto u \end{aligned}$$

with $k \in \mathbb{N}_0$ and $D(F) := D(P)$. The function u is the weak solution of the "re-stated" problem (5.15). If $k \neq 0$ (or in the case of a *time-independent* ρ), this u is also the weak solution of the original wave problem (5.14). This is a consequence of Corollary 3.12, applied to $C(t)v := v/c(t)^2$ and $D(t)v := v/\rho(t)$.

5.2.1 *Injectivity*

We would like to make a short remark about whether the data u uniquely determines the searched for parameter c, ν, q or ρ. We can easily derive conditions under which this will *not* be the case: If $u(t,x) = 0$ in a neighborhood of a point $(t_0, x_0) \in I \times \Omega$, then we can change any of the parameters at this point in space-time without changing the data. It is clear that the time-independent case is much more forgiving. There, we can only give this guarantee if u vanishes in a neighborhood of the line $\{ (t, x_0) \mid t \in I \} \subset I \times \Omega$. In other words, if there are no waves passing through x_0 throughout the whole time interval.

For a more formal investigation of the injectivity of F (with respect to one of the parameters) we can proceed as in our investigation of Lipschitz-continuity (see Lemma 4.3). Let $u_{1,2} := F(c, \nu, q_{1,2}, \rho)$. Since their difference $u_2 - u_1$ is the unique solution of a wave equation with right-hand side $(q_2 - q_1)u_1$, we have $u_2 = u_1 =: u$ if and only if $(q_2(t,x) - q_1(t,x))u(t,x) = 0$ for almost all $(t,x) \in I \times \Omega$. This means that if $u(t,x) \neq 0$ almost everywhere, then $q_2 = q_1$. Similar conditions can be derived for the other parameters. For example, if two wave speeds c_1, c_2 generate the same data u and $((c_2 - c_1)u')' \neq 0$ almost everywhere, then $c_2 = c_1$. However, even $u(t,x) \neq 0$ for almost all (t,x) is hard to satisfy. Due to the homogeneous initial condition $u(0) = 0$ and the finite speed of propagation, this can only be achieved if the right-hand side $f(t)$ is active everywhere in the domain for all $t \in (0, \varepsilon)$ for some $\varepsilon > 0$. The fact that local perturbations of the parameters only cause a local response in the wave field is inherent to wave phenomena [Ika00]. In contrast, even a local change of parameters in a parabolic equation would immediately generate a global reaction in the solution, even if it might decay very quickly with increasing distance (in time or space) from the perturbation.

For a time-dependent q there exist some articles that deal with the injectivity, see [Aïc15; BB19; Esk07; HK18; Kia17; KO17; RS91; RR91; Ste89; Wat14]. The articles [CL05; IS92] deal with a constant q, but also consider the coefficient ν. Due to the aforementioned difficulties, these works treat inverse problems with data that is much richer in information content. Mostly, this is the knowledge about not only one wave field for one source term, but the whole operator that maps sources to measurements. For boundary data, this is usually modeled using a Dirichlet-to-Neumann map or its inverse, depending on the setup.

5.3 FRÉCHET-DIFFERENTIABILITY

We depend on differentiability of F for its numerical inversion using Newton-based methods. The forward operator is comprised of the operators P and S, and the differentiability of S has already been discussed in Section 4.2. Thus, we first need to establish differentiability of the value operator P.

The parameters c and ρ enter into P only by their reciprocal values. Knowing that these are also differentiable is enough to define P, but to prove that P is Fréchet-differentiable (or just continuous) with respect to the norm of $X^{(k)}$, we depend on norm estimates for derivatives of such reciprocal functions. For fixed, small k this is best achieved using straightforward calculation, but even for $k = 3$ this already becomes tedious. Estimates for derivatives of arbitrary order are surprisingly hard to find in the literature, therefore we provide our own through the following Lemma.

Lemma 5.1. *Let* $m \in \mathbb{N}$, $g_0 > 0$ *and* $g \in W^{m,\infty}(I)$ *with* $g(t) \geqslant g_0 > 0$ *almost everywhere. Then*

$$\left\| \frac{1}{g} \right\|_{W^{m,\infty}(I)} \leqslant M \left(1 + g_0^{-1}\right)^{m+1} \left(1 + \|g\|_{W^{m,\infty}(I)}\right)^m$$

holds, where $M > 0$ *is a constant that only depends on* m.

Proof. Clearly, it is sufficient to verify the claim for $g \in C^m(I)$. A formula for derivatives of $1/g$ can be found in [Les91]; for $i = 1, \dots, m$ we have

$$\frac{d^i}{dt^i} \frac{1}{g(t)} = \sum_{k=1}^{i} \frac{(-1)^k}{g(t)^{k+1}} \binom{i+1}{k+1} \frac{d^i}{dt^i} g(t)^k. \tag{5.21}$$

Derivatives of powers of g can be calculated by a general form of the Leibniz rule, which yields

$$\frac{d^i}{dt^i} g(t)^k = \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| = i} \binom{i}{\alpha} \prod_{j=1}^{k} g^{(\alpha_j)}(t).$$

For $k = 1, \dots, m$, the absolute value of $g^{(\alpha_j)}(t)$ is bounded by $\|g\|_{W^{m,\infty}(I)}$, which does not depend on the multiindex α. On the remaining sum we apply the multinomial theorem and obtain

$$\left| \frac{d^i}{dt^i} g(t)^k \right| \leqslant k^i \|g\|^k_{W^{m,\infty}(I)} \leqslant m^m \left(1 + \|g\|_{W^{m,\infty}(I)}\right)^m,$$

which we insert into (5.21) after taking absolute values on both sides. After some coarse estimation, we end up with

$$\begin{aligned} \left| \frac{d^i}{dt^i} \frac{1}{g(t)} \right| &\leqslant \sum_{k=1}^{i} \frac{m^m}{g_0^{k+1}} \binom{i+1}{k+1} \left(1 + \|g\|_{W^{m,\infty}(I)}\right)^m \\ &\leqslant 2^{m+1} m^m \left(1 + g_0^{-1}\right)^{m+1} \left(1 + \|g\|_{W^{m,\infty}(I)}\right)^m \end{aligned}$$

and note that this also holds for $i = 0$. Thus, we have proven the desired estimate with $M := 2^{m+1} m^m$. □

Keeping this lemma in mind, we can turn to proving differentiability of P.

Theorem 5.2. *Let* $k \in \mathbb{N}_0$. *The map* $P\colon D(P) \cap W^{(k)} \to X^{(k)}$ *is Fréchet-differentiable, and its derivative* $\partial P\colon D(P) \cap W^{(k)} \to \mathcal{L}\big(W^{(k)}, X^{(k)}\big)$, *evaluated at* $(c, \nu, \rho, q) \in D(P) \cap W^{(k)}$ *and* $(\bar{c}, \bar{\nu}, \bar{\rho}, \bar{q}) \in W^{(k)}$ *is given by*

$$
\begin{aligned}
\partial P(c, \nu, \rho, q)[\bar{c}, \bar{\nu}, \bar{\rho}, \bar{q}] &= \begin{pmatrix} \partial A_\rho[\bar{\rho}] \\ \partial B_{c,\nu,\rho}[\bar{c}, \bar{\nu}, \bar{\rho}] \\ \partial C_{c,\rho}[\bar{c}, \bar{\rho}] \\ \partial Q_q[\bar{q}] \end{pmatrix} \\
&= t \mapsto \begin{pmatrix} \varphi \in H_0^1 \mapsto \operatorname{div}\left(\dfrac{\bar{\rho}(t)}{\rho(t)^2} \nabla \varphi\right) \\ v \in L^2 \mapsto \left(\dfrac{\bar{\rho}'(t)}{\rho(t)^2 c(t)^2} - \dfrac{2\rho'(t)\bar{\rho}(t)}{\rho(t)^3 c(t)^2} - \dfrac{2\rho'(t)\bar{c}(t)}{\rho(t)^2 c(t)^3} + \bar{\nu}(t)\right) v \\ v \in L^2 \mapsto -\left(\dfrac{\bar{\rho}(t)}{\rho(t)^2 c(t)^2} + \dfrac{2\bar{c}(t)}{\rho(t) c(t)^3}\right) v \\ v \in L^2 \mapsto \bar{q}(t) v \end{pmatrix}.
\end{aligned}
$$

Moreover, $\partial P\colon D(P) \cap W^{(k)} \to \mathcal{L}\big(W^{(k)}, X^{(k)}\big)$ *is continuous.*

Proof. The image of P consists of finitely many components, therefore it is enough to look at each component individually. The proposed candidates for the derivatives of each of the operators A, B, C and Q with respect to the parameters can be obtained by formally treating them as if they were ordinary rational functions with scalar arguments c, ν, ρ and q; for B we only have to note that differentiation in time is a linear operation. The linearity of the resulting operators is obvious, and their boundedness is also easy to see: For instance, time derivatives of $\partial A_\rho[\bar{\rho}]$ of order $i = 0, \ldots, k+1$ can be bounded by

$$
\begin{aligned}
\left| \left\langle (\partial A_\rho[\bar{\rho}])^{(i)}(t) \varphi, \psi \right\rangle \right| &\leqslant \int_\Omega \left| \left(\bar{\rho}/\rho^2\right)^{(i)}(t) \right| \, |\nabla \varphi \cdot \nabla \psi| \, dx \\
&\leqslant \left\| \left(\bar{\rho}/\rho^2\right)^{(i)}(t) \right\|_{L^\infty(\Omega)} \|\varphi\|_{H_0^1(\Omega)} \|\psi\|_{H_0^1(\Omega)} \\
&\leqslant 2^i \|\bar{\rho}\|_{W^{i,\infty}(I;L^\infty(\Omega))} \left\|\rho^{-2}\right\|_{W^{i,\infty}(I;L^\infty(\Omega))} \|\varphi\|_{H_0^1(\Omega)} \|\psi\|_{H_0^1(\Omega)}
\end{aligned}
$$

by making use of the Leibniz rule. To conclude boundedness of the linear map

$$
\partial A_\rho\colon W^{k+1,\infty}(I; L^\infty(\Omega)) \to W^{k+1,\infty}\big(I; \mathcal{L}\big(H_0^1(\Omega), H^{-1}(\Omega)\big)\big)
$$

we can leave the norm of ρ^{-2} as-is, because we only need it to be finite. However, applying Lemma 5.1 to it yields the continuity of ∂A_ρ in ρ, and therefore also in the whole tuple (c, ν, ρ, q) in the norm of $W^{(k)}$.

We demonstrate the estimation of the linearization error in the context of the third component of ∂P, the operator C. For this we need to assume that $\bar{c}$ and $\bar{\rho}$ are small enough such that $(c+\bar{c}, \nu, \rho+\bar{\rho}, q)$ belongs to the open set $D(P) \cap X^{(k)}$. Then we calculate

$$\begin{aligned}
e &:= \|C_{c+\bar{c},\rho+\bar{\rho}} - C_{c,\rho} - \partial C_{c,\rho}[\bar{c},\bar{\rho}]\|_{W^{k+1,\infty}(I;\mathcal{L}(L^2(\Omega)))} \\
&= \left\| \frac{1}{(\rho+\bar{\rho})(c+\bar{c})^2} - \frac{1}{\rho c^2} + \frac{\bar{\rho}}{\rho^2 c^2} + \frac{2\bar{c}}{\rho c^3} \right\|_{W^{k+1,\infty}(I;L^\infty(\Omega))} \\
&= \left\| \frac{\bar{\rho}^2}{c^2\rho^2(\rho+\bar{\rho})} + \frac{3\bar{c}^2 c + 2\bar{c}^3}{c^3(c+\bar{c})^2(\rho+\bar{\rho})} + \frac{2\bar{c}\bar{\rho}}{c^3\rho(\rho+\bar{\rho})} \right\|_{W^{k+1,\infty}(I;L^\infty(\Omega))}.
\end{aligned}$$

Again, we are required to bound not only this difference with respect to $L^\infty(I; L^\infty(\Omega))$ but also its time derivatives. On each fraction we can invoke the product rule, and Lemma 5.1 shows the norms of the denominators to remain bounded when $(\bar{c}, \bar{\rho}) \to 0$. We observe that the $W^{k+1,\infty}(I; L^\infty(\Omega))$-norms of the numerators are of order $\mathcal{O}(\|(\bar{c}, \bar{\rho})\|^2)$, thus the linearization error e has to be as well. □

We can combine this result with Corollary 4.8 to obtain total differentiability of F. However, we know from the previous chapter that differentiability with respect to B or Q can be obtained in a less regular setting. Therefore we will start with partial differentiability of F with respect to q and ν and then go on to show its total differentiability.

Theorem 5.3. *Let $k \geq 0$ and $f \in \mathcal{F}^{(k)}$. Then $F\colon D(P) \cap W^{(k)} \to Y^{(k)}$ is partially Fréchet-differentiable with respect to the variables ν and q. For every $x = (c, \nu, \rho, q) \in D(P) \cap W^{(k)}$ and $h \in W^{k,\infty}(I; L^\infty(\Omega))$, the values $\partial_\nu F(x)[h]$ and $\partial_q F(x)[h]$ are given as the unique weak solutions $u_h \in Y^{(k)}$ of*

$$\left(\frac{u_h'}{\rho c^2}\right)' + \left(\nu + \frac{\rho'}{\rho^2 c^2}\right) u_h' - \operatorname{div}\left(\frac{\nabla u_h}{\rho}\right) + q u_h = g(u)[h],$$

together with homogeneous initial values $u_h(0) = (u_h'/(\rho c^2))(0) = 0$. Here, $u := F(x)$ and the right-hand side for $\partial_\nu F(x)[h]$ reads $g(u)[h] := -hu'$, whereas for $\partial_q F(x)[h]$ it is $g(u)[h] := -hu$.

Proof. From Theorem 4.4 we know about the partial derivatives of the operator $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$ (the solution operator to (5.17)) with respect to the operators B and Q. The parameters ν and q influence exactly one of these operators, thus the derivative of F with respect to q can be written as

$$\partial_q F(c, \nu, \rho, q) = \partial_Q S(P(c, \nu, \rho, q)) \circ \partial_q Q_q.$$

The assertion is obtained by substituting $\partial_Q S(P(c, \nu, \rho, q))$ with the characterization provided by Theorem 4.4. The operator Q is linearly dependent on q, but we still need to use Theorem 5.2 to obtain its boundedness. The same reasoning applies to $\partial_\nu F$. □

If we set k to be zero, we see that $\nu \in W^{1,\infty}(I; L^\infty(\Omega))$ is enough to be able to differentiate F in this parameter. Moreover, we see that the potential q does not need to be equipped with any differentiability in time; $q \in L^\infty(I; L^\infty(\Omega))$ is sufficient. This is an improvement of our earlier results, at least in the time variable. In [GL17] we also derived differentiability for S with respect to q, but in the somewhat easier case $c = \rho = 1$, $\nu = 0$. There, we proved that there is an open subset of $W^{1,\infty}(I; L^2(\Omega))$ on which one can obtain a Fréchet-differentiable S.

We would like to emphasize that the existence of all partial derivatives of S (as we have shown in Theorem 4.4) is not sufficient to obtain differentiability of F with respect to c or ρ, because these parameters influence multiple operators. Instead, we must resort to Corollary 4.8, which shows total differentiability of S, as long as $k \geqslant 2$. Combining this with the differentiability of P, we can also deduce that the whole operator F is Fréchet-differentiable, as the following theorem demonstrates.

Theorem 5.4. *Let $k \geqslant 2$ and $f \in \mathcal{F}^{(k)}$. Then $F\colon D(P) \cap W^{(k)} \to Y^{(k-2)}$ is Fréchet-differentiable. For every $x = (c, \nu, \rho, q) \in D(P) \cap W^{(k)}$ and $h = (\bar{c}, \bar{\nu}, \bar{\rho}, \bar{q}) \in W^{(k)}$, the value $\partial F(x)[h]$ is the unique weak solution $u_h \in Y^{(k-1)}$ of the partial differential equation*

$$\begin{aligned} \frac{1}{\rho}\left(\frac{u_h'}{c^2}\right)' + \nu u_h' - \operatorname{div}\left(\frac{\nabla u_h}{\rho}\right) + q u_h \\ = \frac{\bar{\rho}}{\rho^2}\left(\frac{u'}{c^2}\right)' - \operatorname{div}\left(\frac{\bar{\rho}}{\rho^2}\nabla u\right) + \frac{2}{\rho}\left(\frac{\bar{c}}{c^3}u'\right)' - \bar{q}u - \bar{\nu}u' \end{aligned}$$

together with homogeneous initial values $u_h(0) = u_h'(0) = 0$. As always, $u = F(x)$ denotes the solution of the forward problem.

Proof. In essence, we proceed like in the previous theorem: Since $F = S \circ P$, we have

$$\partial F(x) = \partial S(P(x)) \circ \partial P(x),$$

which can be evaluated at $h \in W^{(k)}$ using Corollary 4.8 and Theorem 5.2. This lets us conclude that $u_h := \partial F(x)[h] \in Y^{(k-1)}$ and solves

$$\begin{aligned} \left(\frac{u_h'}{\rho c^2}\right)' + \left(\nu + \frac{\rho'}{\rho^2 c^2}\right)u_h' - \operatorname{div}\left(\frac{\nabla u_h}{\rho}\right) + q u_h \\ = -\operatorname{div}\left(\frac{\bar{\rho}}{\rho^2}\nabla u\right) - \bar{q}u + \left(\left(\frac{\bar{\rho}}{\rho^2 c^2} + \frac{2\bar{c}}{\rho c^3}\right)u'\right)' \\ - \left(\frac{\bar{\rho}'}{\rho^2 c^2} - \frac{2\rho'\bar{\rho}}{\rho^3 c^2} - \frac{2\rho'\bar{c}}{\rho^2 c^3} + \bar{\nu}\right)u'. \end{aligned}$$

Both sides of this equation can be simplified because both u and u_h belong to $Y^{(1)} = W^{2,\infty}(I; L^2(\Omega)) \cap W^{1,\infty}(I; H_0^1(\Omega)) \subset C^1(I; L^2(\Omega))$. By means of the product rule we obtain the PDE in the assertion. The same holds for the homogeneous initial values of u_h: Because u_h' is continuous (with values in $L^2(\Omega)$), the second initial condition $(u_h'/(\rho c^2))(0) = 0$ is equivalent to $u_h'(0) = 0$. □

We would like to remark that the PDE for ∂F is exactly what one would expect, and that it can also be deduced through formal linearization of the wave equation (5.15). The significance of our result consists of showing that (especially in the context of time-dependent parameters) this linearized equation is well-posed and that it actually is the Fréchet-derivative of F.

5.3.1 *Adjoint of the derivative*

The adjoint of $\partial S(p)$ has been analyzed in Section 4.2.3, motivated by the fact that most regularization methods make use of this adjoint. There, we have also discussed the fact that the space $Y^{(k)}$ is probably too regular for measurement noise, thus we only considered the derivative as the linear map $\partial S(p) \in \mathcal{L}(X^{(k)}, L^2(I; H))$. We continue this investigation with the operator F, and are particularly interested in the adjoint of $\partial F(x) \in \mathcal{L}\big(W^{(k)}, L^2(I; L^2(\Omega))\big)$. From Corollary 4.10 we already know a lot about $\partial S(P(x))^*$, and due to the chain rule we have

$$\partial F(x)^* = \partial P(x)^* \circ \partial S(P(x))^*. \tag{5.22}$$

By identifying the dual space of $L^2(I; L^2(\Omega))$ with the space itself we can regard the adjoint as

$$\partial F(x)^* \in \mathcal{L}\Big(L^2(I; H), \, \big(W^{(k)}\big)^*\Big).$$

Equation (5.22) suggests that we should also analyze $\partial P(x)^*$ independently from S. However, even with the simple structure of P (most of its components are simple multiplication operators) a characterization of $\partial P(x)^* \in \mathcal{L}\Big(\big(X^{(k)}\big)^*, \big(W^{(k)}\big)^*\Big)$ is not possible due to our insufficient knowledge about the dual space of $X^{(k)}$. We simply have no idea how a general $v \in (X^{(k)})^*$ could act on $\partial P(x)[h]$. This is not only true for P as a whole but also for its simplest components: Even in the case $k = 0$, finding a closed formula for the adjoint of

$$Q \colon L^\infty(I; L^\infty(\Omega)) \to L^\infty(I; \mathcal{L}(L^2(\Omega))), \quad q \mapsto Q_q$$

does not seem to be possible.

Fortunately, we do not need to evaluate $P^*(z)$ for arbitrary z, but only for $z \in \mathcal{R}(\partial S(P(x))^*)$. From Corollary 4.10 we know that these $z \in \big(X^{(k)}\big)^*$ evaluate their arguments at a point (that depends on x) and form a kind of $L^2(I; L^2(\Omega))$ inner product with the results.

Theorem 5.5. *Let* $k \geqslant 2$, $f \in \mathcal{F}^{(k)}$ *and* $x = (c, \nu, \rho, q) \in D(P) \cap W^{(k)}$. *The application of the adjoint of* $\partial F(x) \in \mathcal{L}(W^{(k)}, L^2(I; L^2(\Omega)))$ *on* $v \in L^2(I; L^2(\Omega))$ *can be written as*

$$\partial F(x)^*[v] = \begin{pmatrix} \dfrac{2u'}{c^3}\left(\dfrac{w_v}{\rho}\right)' \\ -u' w_v \\ \dfrac{\nabla u \cdot \nabla w_v}{\rho^2} + \dfrac{w_v}{\rho^2}\left(\dfrac{u'}{c^2}\right)' \\ -u w_v \end{pmatrix} \in L^\infty(I; L^1(\Omega))^4, \tag{5.23}$$

where the embedding of $L^\infty(I;L^1(\Omega))$ *into* $(W^{(k)})^*$ *has to be understood using the inner product of* $L^2(I;L^2(\Omega))$ *and* $u = F(x) \in Y^{(2)}$. *With* $w_\nu \in Y$ *we denote the solution of the adjoint equation*

$$\left(\frac{1}{c^2}\left(\frac{w_\nu}{\rho}\right)'\right)' - (\gamma w_\nu)' - \operatorname{div}\left(\frac{\nabla w_\nu}{\rho}\right) + q w_\nu = \nu \tag{5.24a}$$

in $L^2(I;H^{-1}(\Omega))$, *together with homogeneous end conditions*

$$w_\nu(T) = \left(\frac{w_\nu}{\rho c^2}\right)'(T) = 0. \tag{5.24b}$$

Proof. Let $h = (\bar{c}, \bar{\nu}, \bar{\rho}, \bar{q}) \in W^{(k)}$. To calculate $\langle \partial F(x)^*[\nu], h\rangle$, we substitute ∂P using Theorem 5.4, and $\partial S(P(x))^*$ using Corollary 4.10 and obtain

$$\begin{aligned}
&\langle \partial F(x)^*[\nu], h\rangle_{(W^{(k)})^*\times W^{(k)}} = \langle \partial S(P(x))^*[\nu], \partial P(x)[h]\rangle_{(X^{(k)})^*\times X^{(k)}} \\
&= \int_0^T \big(\partial C_{c,\rho}[\bar{c},\bar{\rho}](t)u'(t), w_\nu'(t)\big) - \big(\partial B_{c,\nu,\rho}[\bar{c},\bar{\nu},\bar{\rho}](t)u'(t), w_\nu(t)\big) \\
&\qquad - \langle (\partial A_\rho[\bar{\rho}](t) + Q_q[\bar{q}](t))u(t), w_\nu(t)\rangle \, dt \\
&= \int_0^T \left(\left(\frac{\bar{\rho}(t)}{\rho(t)^2 c(t)^2} + \frac{2\bar{c}(t)}{\rho(t)c(t)^3}\right)u'(t),\, w_\nu'(t)\right) \\
&\qquad - \left(\left(\frac{\bar{\rho}'(t)}{\rho(t)^2 c(t)^2} - \frac{2\rho'(t)\bar{\rho}(t)}{\rho(t)^3 c(t)^2} - \frac{2\rho'(t)\bar{c}(t)}{\rho(t)^2 c(t)^3} + \bar{\nu}(t)\right)u'(t),\, w_\nu(t)\right) \\
&\qquad - \left\langle \operatorname{div}\left(\frac{\bar{\rho}(t)}{\rho(t)^2}\nabla u(t)\right) + \bar{q}(t)u(t),\, w_\nu(t)\right\rangle dt,
\end{aligned}$$

where w_ν denotes the solution of the adjoint problem (4.14) on page 58. Our goal is to reshape this expression into some kind of dual product that has h on one side. Inside all of the $L^2(\Omega)$ inner products we can shift from one side to the other as we wish, as long as both sides of the inner product belong to $L^2(\Omega)$. In general, neither the product $u(t)w_\nu(t)$ of two $H_0^1(\Omega)$-functions, nor the multiplication of $u'(t) \in H_0^1(\Omega)$ and $w_\nu'(t) \in L^2(\Omega)$ will belong to $L^2(\Omega)$. However, they do lie in $L^1(\Omega)$. Thus, we resort to interpreting the resulting integrals as dual products between $L^\infty(\Omega)$ and $L^1(\Omega)$. This and some further reorganizing yields

$$\begin{aligned}
\langle \partial F(x)^*[\nu], h\rangle = \int_0^T &\left\langle \frac{2u'(t)}{c(t)^3}\frac{d}{dt}\left(\frac{w_\nu(t)}{\rho(t)}\right), \bar{c}(t)\right\rangle_{L^\infty\times L^1} \\
&+ \left\langle \frac{\nabla u(t)\cdot\nabla w_\nu(t)}{\rho(t)^2} + \frac{w_\nu(t)}{\rho(t)^2}\frac{d}{dt}\left(\frac{u'(t)}{c(t)^2}\right), \bar{\rho}(t)\right\rangle_{L^\infty\times L^1} \\
&- \langle u'(t)w_\nu(t), \bar{\nu}(t)\rangle_{L^\infty\times L^1} \\
&- \langle u(t)w_\nu(t), \bar{q}(t)\rangle_{L^\infty\times L^1} \, dt.
\end{aligned} \tag{5.25}$$

This proves that the adjoint has the asserted form, with w_ν as the solution of (4.14), which reads (keeping in mind that Q_q and $B_{c,\nu,\rho}$ are self-adjoint)

$$(\mathcal{C}_{c,\rho} w_\nu')' - \mathcal{B}_{c,\nu,\rho} w_\nu' + (\mathcal{A}_\rho + \mathcal{Q}_q - \mathcal{B}_{c,\nu,\rho}') w_\nu = \nu. \tag{5.26}$$

This equation has to hold in the $L^2(I; H^{-1}(\Omega))$-sense, and is joined by the end conditions $w_\nu(T) = (\mathcal{C}_{c,\rho} w_\nu')(T) = 0$, which directly translate to those given in the assertion. Substituting the operators with their definitions, equation (5.26) becomes

$$\left(\frac{w_\nu'}{\rho c^2}\right)' - \left(\nu + \frac{\rho'}{\rho^2 c^2}\right) w_\nu' - \left(\nu + \frac{\rho'}{\rho^2 c^2}\right)' w_\nu - \operatorname{div}\left(\frac{\nabla w_\nu}{\rho}\right) + q w_\nu = \nu.$$

In this form we can handle the PDE with the theory developed in Chapters 2 and 3, but two applications of the product rule prove that it is equivalent to the more compact differential equation (5.24a). □

We wish to make two remarks. First, as consequence of Theorem 5.3 and Theorem 4.9, the result also holds for $k \geqslant 0$ if we restrict the derivatives under consideration to be $\partial_q F$ and $\partial_\nu F$. Second, $\partial F(x)^*[v]$ (as in (5.23)) will not only belong to $L^\infty(I; L^1(\Omega))^4$. For $k \geqslant 2$ we have $u \in W^{2,\infty}(I; H_0^1(\Omega))$, $w_\nu \in L^\infty(I; H_0^1(\Omega))$ and $w_\nu' \in L^\infty(I; L^2(\Omega))$, and the embedding theorems for Sobolev spaces (cf. [AF03]) yield $H_0^1(\Omega) \subset L^p(\Omega)$ for $p > 2$, but this depends on the space dimension d. Combining this with the Hölder-inequality, we see that in the case of a one-, two- or three-dimensional problem the following holds:

- If $d = 1$, then $H_0^1(\Omega) \subset L^\infty(\Omega)$ and
$$(\partial F(x)^*[v])(t) \in L^2(\Omega) \times L^\infty(\Omega) \times L^1(\Omega) \times L^\infty(\Omega).$$
- If $d = 2$, then $H_0^1(\Omega) \subset L^p(\Omega)$ for all $1 \leqslant p < \infty$ and
$$(\partial F(x)^*[v])(t) \in L^{q_1}(\Omega) \times L^{q_2}(\Omega) \times L^1(\Omega) \times L^{q_2}(\Omega)$$
for all $1 \leqslant q_1 < 2$ and $1 \leqslant q_2 < \infty$.
- If $d = 3$, then $H_0^1(\Omega) \subset L^p(\Omega)$ for all $1 \leqslant p < 6$ and
$$(\partial F(x)^*[v])(t) \in L^{q_1}(\Omega) \times L^{q_2}(\Omega) \times L^1(\Omega) \times L^{q_2}(\Omega)$$
for all $1 \leqslant q_1 < 3/2$ and $1 \leqslant q_2 < 3$.

Note that we cannot use regularity results for w_ν because this would require more than just $v \in L^2(I; L^2(\Omega))$, which means we would have to change the space of the measurements. We see that we can *almost* obtain a setting for the adjoint that only involves *reflexive* Banach spaces. However, for the third component of $\partial F(x)^*[v]$ (which contains $\nabla u \cdot \nabla w_\nu$) to belong to something else than $L^1(\Omega)$, we would first need to prove better *spatial* regularity results for u or w_ν.

5.4 ILL-POSEDNESS

In Theorem 4.13 we showed the ill-posedness of S by constructing suitable sequences of arguments. These sequences do not lie in the range of P, therefore we cannot use that result to directly conclude ill-posedness of F. Instead, we construct sequences of parameters such that their images under P fulfill the assumptions of Theorem 4.11.

In the abstract operator setting we used time-independent disturbances. This time we decide to make them independent of the spatial variables instead. Working in the time variable is more difficult, because the parameters (and hence also the perturbations) have to be smooth in time. The following Lemma provides smooth auxiliary functions, that we will subsequently use to construct suitable sequences of parameters.

Lemma 5.6. *Let $r \in \mathbb{N} \cup \{0\}$. There exists $(\alpha_j)_{j\in\mathbb{N}} \subset C_c^\infty(I)$ which satisfies*

$$0 < \gamma \leqslant \left\|\alpha_j\right\|_{W^{r,\infty}(I)} \leqslant 1 \quad \textit{for all } j \in \mathbb{N}$$

and $\alpha_j \varphi \to 0$ in $H^m(I)$ as $j \to \infty$ for all fixed $\varphi \in H^m(I)$ with $m = 0, \ldots, r$. Moreover, if $r \geqslant 1$ then $\left\|\alpha_j\right\|_{W^{r-1,\infty}(I)} \to 0$ when $j \to \infty$.

Proof. Let $t_0 \in (0, T)$ and $\psi \in C_c^\infty(\mathbb{R})$ with $\operatorname{spt}\psi = [-1, 1]$, $\psi(t) \in [0, 1]$ for all $t \in \mathbb{R}$ and $\|\psi\|_{W^{r,\infty}(I)} = 1$. We define

$$\alpha_j(t) = j^{-r}\psi(j(t - t_0)).$$

If $j > \max\left\{t_0^{-1}, (T - t_0)^{-1}\right\}$, then $\alpha_j \in C_c^\infty(I)$. Hence, we might have to discard some elements from the beginning of this sequence, but without loss of generality we assume that this is not the case. We see that $\operatorname{spt}\alpha_j = t_0 + [-1/j, 1/j]$ and $\alpha_j^{(i)}(t) = j^{i-r}\psi^{(i)}(j(t - t_0))$. Thus,

$$\left\|\alpha_j\right\|_{W^{r,\infty}(I)} = \max_{i=0,\ldots,r} j^{i-r}\left\|\psi^{(i)}\right\|_{L^\infty(I)}$$

is less than 1 and greater than $\left\|\psi^{(r)}\right\|_{L^\infty(I)}$. We also see from this that $\left\|\alpha_j\right\|_{W^{r-1,\infty}(I)}$ tends to zero when $j \to \infty$ and $r > 0$. For arbitrary $\varphi \in H^m(I)$ with $m \in \{0, \ldots, r\}$, we deduce that

$$\begin{aligned}
\left\|\alpha_j\varphi\right\|_{H^m(I)}^2 &= \sum_{i=0}^{m}\sum_{l=0}^{i}\binom{i}{l}\int_0^T\left|\alpha_j^{(i-l)}(t)\varphi^{(l)}(t)\right|^2 dt \\
&\leqslant \sum_{i=0}^{m}\sum_{l=0}^{i}\binom{i}{l}\int_{t_0-1/j}^{t_0+1/j}\left|\varphi^{(l)}(t)\right|^2 dt \to 0,
\end{aligned}$$

because the integrand $\mathbb{1}_{[t_0-1/j, t_0+1/j]}(t)\,|\varphi^{(l)}(t)|^2$ converges pointwise to zero and is dominated by the integrable function $|\varphi^{(l)}|^2$. □

In the proof of the following result we use these functions to construct appropriate sequences of perturbations.

Theorem 5.7. *Let $k \in \mathbb{N} \cup \{0\}$, $f \in \mathcal{F}^{(k)}$ and $p = (c, \nu, \rho, q) \in D(P) \cap W^{(k)}$.*

(i) The inverse problems of finding v *or* q *such that* $F(p) = y \in Y^{(k)}$ *are locally ill-posed.*

(ii) If $k \in \mathbb{N}$*, then the tasks of finding* c *or* ρ *such that* $F(p) = y \in Y^{(k-1)}$ *holds are also locally ill-posed.*

Note that the ill-posedness of determining a time-dependent q was already proven in [GL17], albeit for different function spaces. Furthermore, assertion (ii) in the theorem generalizes the results of [KR14b]. There, it was shown that a time-independent wave speed and mass density already cause the problems to be locally ill-posed. By choosing the sequences as purely time-dependent our result shows that an entirely dynamic wave speed and density have the same effect.

Proof. Let $p = (c, v, \rho, q) \in D(P) \cap W^{(k)}$ and $(A, B, C, Q) := P(p)$. Since $D(P) \cap W^{(k)}$ forms an open subset of $W^{(k)}$, there exists $\delta_0 > 0$ such that $B(p, \delta_0) \subset D(P) \cap W^{(k)}$. Let $0 < \delta \leqslant \delta_0$.

(i) Reconstruction of q: With $(\alpha_j)_{j \in \mathbb{N}}$ we denote the sequence from Lemma 5.6 for $r = k$. We use it to define $q_j(t, x) := q(t, x) + \delta\alpha_j(t)$ for $(t, x) \in I \times \Omega$. By design, $q_j \not\to q$ in $W^{k,\infty}(I; L^\infty(\Omega))$, $q_j \in B(q, \delta)$ and we have $P(c, v, \rho, q_j) = (A, B, C, Q + R_j)$ with R_j given as

$$R_j \in W^{k,\infty}\big(I; \mathcal{L}\big(L^2(\Omega)\big)\big), \quad R_j(t)\nu = \delta\alpha_j(t)\nu.$$

The properties of α_j imply $\mathcal{R}_j\nu \to 0$ with respect to the $H^k(I; L^2(\Omega))$-norm for all $\nu \in H^k(I; L^2(\Omega))$. Thus, we can apply assertion (i) of Theorem 4.11 to conclude

$$F(c, v, \rho, q_j) = S(A, B, C, Q + R_j) \to S(A, B, C, Q) = F(p)$$

in the norm of $Y^{(k)}$ as $j \to \infty$.

Reconstruction of v: We set $k_1 := \max\{1, k\}$. The argument is similar to q, but we have to define $(\alpha_j)_{j \in \mathbb{N}}$ to be the sequence that we obtain from Lemma 5.6 after setting $r = k_1$. From this we build $v_j(t, x) := v(t, x) + \delta\alpha_j(t)$ for $(t, x) \in I \times \Omega$. Again, $v_j \not\to v$ in $W^{k_1,\infty}(I; L^\infty(\Omega))$ as $j \to \infty$, $v_j \in B(v, \delta)$ and $P(c, v_j, \rho, q) = (A, B + R_j, C, Q)$ with

$$R_j \in W^{k_1,\infty}\big(I; \mathcal{L}\big(L^2(\Omega)\big)\big), \quad R_j(t)\nu = \delta\alpha_j(t)\nu.$$

We see that $\mathcal{R}_j\nu \to 0$ in $H^k(I; L^2(\Omega))$ for every $\nu \in H^k(I; L^2(\Omega))$, thus in particular $\mathcal{R}_j\nu' \to 0$ for all $\nu \in H^{k+1}(I; L^2(\Omega)) \supset Y^{(k)}$. From result (ii) of Theorem 4.11 follows $F(c, v_j, \rho, q) = S(A, B + R_j, C, Q) \to S(A, B, C, Q) = F(p)$ in $Y^{(k)}$ when $j \to \infty$.

(ii) Reconstruction of ρ: The parameter ρ is involved in three operators, making this part of the proof more complicated. Moreover, we have to take care that the perturbations still satisfy the coercivity constraints. Let α_j again denote the sequence for $r = k + 1$. We define

$$\rho_j(t, x) := \frac{1}{\varepsilon\alpha_j(t) + \rho(t, x)^{-1}}$$

and $p_j := (c, \nu, \rho_j, q)$, where we would like to set $\varepsilon > 0$ in such a way that $\rho_j \in B(\rho, \delta)$ for all $j \in \mathbb{N}$. To make ρ_j well-defined, we require $\varepsilon < 1/\|\rho\|_{L^\infty(I;L^\infty(\Omega))}$, and to secure some wiggle room we further restrict this to $\varepsilon \leqslant 1/(2\rho_1)$. This implies $\varepsilon\alpha_j + \rho^{-1} \geqslant 1/(2\rho_1)$. Denoting with M the positive constant from Lemma 5.1 for $m = k+1$, we see that

$$
\begin{aligned}
\|\rho_j - \rho\|_{W^{k+1,\infty}(I;L^\infty(\Omega))} &= \left\| \frac{\varepsilon\alpha_j\rho}{\varepsilon\alpha_j + \rho^{-1}} \right\|_{W^{k+1,\infty}(I;L^\infty(\Omega))} \\
&\leqslant 2^{k+1}\varepsilon\|\alpha_j\|_{W^{k+1,\infty}(I;L^\infty)} \left\| \frac{\rho}{\varepsilon\alpha_j + \rho^{-1}} \right\|_{W^{k+1,\infty}(I;L^\infty)} \\
&\leqslant 4^{k+1}\varepsilon\|\alpha_j\|_{W^{k+1,\infty}(I;L^\infty)}\|\rho\|_{W^{k+1,\infty}(I;L^\infty)} \left\|(\varepsilon\alpha_j + \rho^{-1})^{-1}\right\|_{W^{k+1,\infty}(I;L^\infty)} \\
&\leqslant 4^{k+1}\varepsilon M\|\rho\|(1+2\rho_1)^{k+2}\Big(1 + \|\varepsilon\alpha_j + \rho^{-1}\|_{W^{k+1,\infty}(I;L^\infty)}\Big)^{k+1} \\
&\leqslant 4^{k+1}\varepsilon M\|\rho\|(1+2\rho_1)^{k+2}\left(1 + \frac{1}{2\rho_1} + \|\rho^{-1}\|_{W^{k+1,\infty}(I;L^\infty)}\right)^{k+1} =: \varepsilon\Lambda,
\end{aligned}
$$

where we abbreviated the $W^{k+1,\infty}(I;L^\infty(\Omega))$-norm of ρ by $\|\rho\|$. The constant Λ only depends on ρ and k. Thus, by setting $\varepsilon := \min\{\delta/\Lambda, 1/(2\rho_1)\}$ we obtain $\rho_j \in B(\rho, \delta)$ for all $j \in \mathbb{N}$, and in consequence $p_j \in B(p, \delta) \subset D(P)$. We will now verify $\rho_j \not\to \rho$ in $W^{k+1,\infty}(I;L^\infty(\Omega))$ by showing that the derivatives of order $k+1$ do not converge with respect to $L^\infty(I;L^\infty(\Omega))$. Expanding $(\rho_j - \rho)^{(k+1)}$ leads to

$$
\begin{aligned}
(\rho_j - \rho)^{(k+1)} &= \left(\frac{\varepsilon\alpha_j\rho}{\varepsilon\alpha_j + \rho^{-1}}\right)^{(k+1)} \\
&= \varepsilon\sum_{i=0}^{k+1}\binom{k+1}{i}\alpha_j^{(i)}\left(\frac{\rho}{\varepsilon\alpha_j + \rho^{-1}}\right)^{(k+1-i)}.
\end{aligned}
$$

Derivatives of $\rho/(\varepsilon\alpha_j + \rho^{-1})$ remain bounded when $j \to \infty$, and $\alpha_j^{(i)} \to 0$ in $L^\infty(I)$ for all $i = 0, \dots, k$. Hence, all summands except for the last one converge to zero. It is therefore sufficient to show that the last summand does not converge to zero in order to conclude this for the whole sum. Indeed, $\alpha_j^{(k+1)}$ does not vanish in the limit, and we observe that $\rho/(\varepsilon\alpha_j + \rho^{-1}) > \rho_0/(\varepsilon + \rho_0^{-1})$ holds almost everywhere. The only thing left to show is the convergence $F(p_j) \to F(p)$ in $Y^{(k-1)}$. We see that $P(p_j) = (A + R_j^A, B + R_j^B, C + R_j^C, Q)$ with

$$
\begin{aligned}
R_j^A(t)v &= \operatorname{div}\left(\left(\frac{1}{\rho_j(t)} - \frac{1}{\rho(t)}\right)\nabla v\right) = \operatorname{div}\big(\varepsilon\alpha_j(t)\nabla v\big), \\
R_j^B(t)u &= \left(\frac{\rho_j'(t)}{\rho_j(t)^2c(t)^2} - \frac{\rho'(t)}{\rho(t)^2c(t)^2}\right)u \\
&= \frac{d}{dt}\left(\frac{1}{\rho(t)} - \frac{1}{\rho_j(t)}\right)\frac{1}{c(t)^2}u = -\frac{\varepsilon\alpha'(t)}{c(t)^2}u, \\
R_j^C(t)u &= \left(\frac{1}{\rho_j(t)c(t)^2} - \frac{1}{\rho(t)c(t)^2}\right)u = \frac{\varepsilon\alpha_j(t)}{c(t)^2}u
\end{aligned}
$$

for all $v \in H_0^1(\Omega)$, $u \in L^2(\Omega)$ and almost all $t \in I$. Since P is continuous, the norms of R_j^A, R_j^B and R_j^C have to remain bounded when $j \to \infty$. Moreover, for every $u \in H^{k+1}(I; L^2(\Omega))$ we see

$$\left\| (\mathcal{R}_j^C u')' \right\|_{H^{k-1}(I;L^2)} \leqslant \|R_j^C u'\|_{H^k(I;L^2)} = \varepsilon \|\alpha_j u'/c^2\|_{H^k(I;L^2)} \to 0$$

as $j \to \infty$ due to the design of the α_j. Likewise, for all $v \in H^k(I; H_0^1(\Omega))$, $\|\mathcal{R}_j^A v\|_{H^k(I;H^{-1})} = \varepsilon \|\alpha_j \Delta v\|_{H^k(I;H^{-1})}$ has to vanish in the limit, and for all fixed $u \in H^k(I; L^2)$ holds

$$\|\mathcal{R}_j^B u'\|_{H^{k-1}(I;L^2)} = \varepsilon \|\alpha_j' u'/c^2\|_{H^{k-1}(I;L^2)} \leqslant \varepsilon \|\alpha_j u'/c^2\|_{H^k(I;L^2)} \to 0.$$

Due to the uniform convergences in Theorem 4.11, we can apply it simultaneously to multiple components of S and finally conclude this part of the proof with

$$S(A + R_j^A, B + R_j^B, C + R_j^C, Q) \to S(A, B, C, Q) = F(p),$$

holding in $Y^{(k-1)}$ as $j \to \infty$.

For the reconstruction of c we can use almost the same approach as for ρ. However, we must take into account that the square of c enters the PDE. We define

$$c_j(t, x) := \left(\varepsilon \alpha_j(t) + c(t, x)^{-2} \right)^{-1/2}$$

for $(t, x) \in I \times \Omega$, and again see that $\varepsilon > 0$ can be chosen in such a way that $c_j \in B(c, \delta)$ for all $j \in \mathbb{N}$. Moreover, $c_j \not\to c$ in $W^{k+1,\infty}(I; L^\infty(\Omega))$ and $P(c_j, v, \rho, q) = (A, B + R_j^B, C + R_j^C, Q)$ with

$$R_j^B(t)u = \left(\frac{\rho'(t)}{\rho(t)^2 c_j(t)^2} - \frac{\rho'(t)}{\rho(t)^2 c(t)^2} \right) u = \frac{\varepsilon \alpha(t) \rho'(t)}{\rho(t)} u,$$

$$R_j^C(t)u = \left(\frac{1}{\rho(t) c_j(t)^2} - \frac{1}{\rho(t) c(t)^2} \right) u = \frac{\varepsilon \alpha_j(t)}{\rho(t)} u.$$

By design, these operators possess the same properties as those that appeared in the proof for ρ. □

A direct consequence of the theorem is the ill-posedness of $F(p) = y \in Y^{(k-1)}$ in all c, v, ρ or q such that $p = (c, v, \rho, q) \in D(P) \cap W^{(k)}$ with $k \geqslant 1$. Note that this is a stronger result than the local ill-posedness of $F(p) = y \in Y^{(k-1)}$ in all tuples $p \in D(P) \cap W^{(k)}$ because the latter problem would already be ill-posed if finding *one* of the parameters yielded an ill-posed problem.

When applying a Newton solver to the nonlinear inverse problem, it is even more important to know whether the linearization of F is ill-posed.

Corollary 5.8. *Let $k \geqslant 2$ and $f \in \mathcal{F}^{(k)}$. We consider $F\colon D(P) \cap W^{(k)} \to Z$ with $Z = W^{j,p}(I; H)$ or $Z = C^j(I; H)$ for $0 \leqslant j \leqslant k$ and $1 \leqslant p < \infty$. For every $p = (c, v, \rho, q) \in D(P) \cap W^{(k)}$, the operator $\partial F(p) \in \mathcal{L}(W^{(k)}, Z)$ is compact.*

Proof. $\partial F(p) = \partial S(P(p)) \circ \partial P(p)$ with linear and continuous $\partial P(p)$, and we know from Lemma 4.16 that $\partial S(P(p))$ is compact with respect to the image spaces given in the assertion. $\square$

It could be that $\partial F(p)$ is compact because it has finite-dimensional range, which would make the resulting problems well-posed in the sense of linear inverse problems (ill-posed in the sense of Hadamard, but with a continuous generalized inverse). Like in the abstract framework this is not the case here.

Lemma 5.9. *Let* $k \in \mathbb{N}_0$, $f \in \mathcal{F}^{(k)}$ *and* $f \neq 0$. *The range of the following operators is infinite-dimensional for every* $p \in D(P) \cap W^{(k)}$*:*

(i) $\partial_q F(p) \in \mathcal{L}\big(W^{k,\infty}(I; L^\infty(\Omega)), Y^{(k)}\big)$,

(ii) $\partial_\rho F(p) \in \mathcal{L}\big(W^{k+1,\infty}(I; L^\infty(\Omega)), Y^{(k-1)}\big)$ *if* $k \geqslant 2$,

(iii) $\partial_\nu F(p) \in \mathcal{L}\big(W^{k,\infty}(I; L^\infty(\Omega)), Y^{(k)}\big)$ *if* $k \geqslant 1$, *and*

(iv) $\partial_c F(p) \in \mathcal{L}\big(W^{k+1,\infty}(I; L^\infty(\Omega)), Y^{(k-1)}\big)$ *if* $k \geqslant 2$.

Proof. Let $(c, \nu, \rho, q) := p$. The technique is very similar to the one we used for the proof of Lemma 4.17. There we made the argument that if e.g. the set of all possible right-hand sides to $u_h = \partial_Q S(x)[h]$ (with the linearization argument h as the variable parameter) was linearly independent, then so is the set of all u_h and in consequence $\partial_Q S(x)$ must possess a infinite-dimensional range. We check this for each of the parameters.

(i) In the part of the proof of Lemma 4.17 that concerns Q and A we used the sequence $(\beta_i)_{i\in\mathbb{N}} \subset C^\infty(I)$. It was chosen in such a way that the set $\{\beta_i u\}_{i\in\mathbb{N}} \subset C(I; L^2(\Omega))$ is linearly independent. We can also make use of this sequence for the parameter q by defining $h_i(t,x) := \beta_i(t)$ for $(t,x) \in I \times \Omega$, and in doing so obtain $h_i \in W^{k,\infty}(I; L^\infty(\Omega))$. The right-hand side of the partial differential equation that is solved by $\partial_q F(p)[h_i]$ then reads $-\beta_i u$, and the set $\{-\beta_i u\}_{i\in\mathbb{N}}$ is linearly independent.

(ii) For ρ we start the same way as for q, but the right-hand side used by $\partial_\rho F(p)[h_i]$ is more complicated; according to Corollary 4.8 it reads

$$\beta_i \left(\frac{1}{\rho^2} \left(\frac{u'}{c^2} \right)' - \operatorname{div}\left(\frac{\nabla u}{\rho^2} \right) \right) =: \beta_i w.$$

Clearly, $w \in C^1(I; H^{-1}(\Omega))$ because $k \geqslant 2$. Moreover, if $w = 0$ then u would solve the wave equation

$$\frac{1}{\rho^2} \left(\frac{u'}{c^2} \right)' - \operatorname{div}\left(\frac{\nabla u}{\rho^2} \right) = 0$$

with homogeneous initial- and boundary conditions. The energy estimates from Section 2.2 state that in this case $u = 0$, which cannot be, because it in turn would imply $f = 0$. Since w is continuous in time and does not vanish everywhere, we can choose β_i in such a way that $\{\beta_i w\}_{i\in\mathbb{N}}$ is linearly independent.

(iii) For v we have to require $k \geqslant 1$ in order to have $u' \in C(I; L^2(\Omega))$. Further, $u' \neq 0$ since $u(0) = 0$ and $u \neq 0$. Hence, we may choose $\beta_i \in C^\infty(I)$ with linearly independent $\{\beta_i u'\}_{i \in \mathbb{N}}$ and proceed in the same way as for q.

(iv) In the derivative with respect to c we can also use this sequence. The right-hand sides $(\beta_i u'/c^3)' = \beta_i' u'/c^3 + \beta_i (u'/c^3)'$ have pairwise disjoint supports inside $\operatorname{spt} u'$ (which is the same as $\operatorname{spt} u'/c^3$), thus they must be linearly independent as long as they do not vanish everywhere. However, it could be that $(\beta_i u'/c^3)' = 0$ for some $i \in \mathbb{N}$. We show how this can be remedied. Suppose this is the case and fix some $t_0 \in I$ and $\varepsilon > 0$ such that $\beta_i'(t_0) \neq 0$, $u'(t_0) \neq 0$ and $\operatorname{spt} \beta_i \subset (t_0 - \varepsilon, t_0 + \varepsilon) \subset I$. We replace β_i with its mirrored version

$$\bar{\beta}_i(t) \coloneqq \begin{cases} \beta_i(2t_0 - t) & \text{if } t \in (t_0 - \varepsilon, t_0 + \varepsilon), \\ 0 & \text{otherwise,} \end{cases}$$

and conclude the proof by observing that

$$\begin{aligned} \left(\bar{\beta}_i \frac{u'}{c^3} \right)'(t_0) &= -\beta_i'(t_0) \frac{u'(t_0)}{c(t_0)^3} + \beta_i(t_0) \left(\frac{u'}{c^3} \right)'(t_0) \\ &= \left(\beta_i \frac{u'}{c^3} \right)'(t_0) - 2\beta_i'(t_0) \frac{u'(t_0)}{c(t_0)^3} = -2\beta_i'(t_0) \frac{u'(t_0)}{c(t_0)^3} \neq 0. \qquad \square \end{aligned}$$

6 APPLICATION TO OTHER EQUATIONS

The main motivation of this work is the acoustic wave equation, and Chapter 4's abstract framework was mainly intended to be applied to this equation, as demonstrated in the last chapter. Nevertheless, we split the theory into a general, evolution equation based part and a part that is specific to the underlying partial differential equation, and we theorized that this general framework could be used for other equations as well. We have not yet proven this fact, but we will do so in this chapter using simplified example problems from linear elasticity and electrodynamics. These equations will not contain as many parameters as we allowed in the wave equation and the overall approach is similar to Chapter 5. Thus, we will keep the presentation short and focus on problem-specific differences to last chapter's results and proofs.

Most of the contents of this chapter can also be found in [Ger19].

6.1 LINEAR ELASTICITY

As a first additional example we would like to consider the propagation of elastic waves through a bounded Lipschitz domain $\Omega \subset \mathbb{R}^3$ in the finite time interval $I := [0, T]$ with $T > 0$. Our model for the displacement field $u\colon I \times \Omega \to \mathbb{R}^3$ is given through the equation

$$(\rho u')' = \operatorname{div} \sigma(u) + f \quad \text{in } (0, T) \times \Omega. \tag{6.1}$$

The right-hand side contains the *restoring force*, which is equal to the row-wise divergence of the stress tensor $\sigma(u)\colon I \times \Omega \to \mathbb{R}^{3\times 3}$ according to Hooke's law. We also allow for a volumetric force $f\colon I \times \Omega \to \mathbb{R}$. The function ρ denotes the mass density inside Ω.

We assume that Ω consists of a linear isotropic material such that the stress tensor is a linear operator and has the form

$$\sigma_{\lambda,\mu} u = 2\mu\varepsilon(u) + \lambda \operatorname{div}(u) I_3.$$

The functions λ and μ denote the space- and time-dependent Lamé coefficients of the material, and I_3 is the 3×3 unit matrix. The symmetric strain tensor $\varepsilon(u) = (Du + Du^\top)/2$ depends on the Jacobian (acting only on the spatial variables) Du of u.

As in Chapter 5, we make the assumption that the material is at rest at $t = 0$, i.e. $u(0) = u'(0) = 0$, and that the body is fixed throughout the whole time interval. This is modeled by the homogeneous Dirichlet boundary condition $u = 0$ on $(0, T) \times \partial\Omega$. This setup implies that the excitation of waves inside Ω happens only due to the volumetric force f.

The inverse problem we would like to consider in this section is the identification of the density ρ and the Lamé coefficients λ, μ from the displacement field u. This setting is relevant for nondestructive testing, where a deviation in these values might indicate a defect in the material.

The reconstruction of space-dependent Lamé coefficients $\lambda, \mu \in L^\infty(\Omega)$ in (6.1) has been analyzed by Lechleiter and Schlasche in [LS17]. The results we obtain in this section are actually quite similar to their results, but our abstract theory allows for time-dependent parameters and shorter presentation. The work of Kirsch and Rieder [KR16] additionally permits the identification of a space-dependent mass density $\rho \in L^\infty(\Omega)$ in the context of a first-order system, and also proves ill-posedness.

The abstract theory presented in Chapter 4 allows to generalize these existing results to time- and space-dependent λ, μ and ρ, because the weak formulation of (6.1) can be written as an evolution equation with the form (4.1). We only indirectly seek to identify the operators in that equation. As before, we will have to compose the solution operator of the abstract equation with a value operator P that maps our three parameters onto suitable linear maps. Computing the Fréchet-derivative of the forward operator is then even easier than for the acoustic wave equation, because P will turn out to be linear.

Since the identification of space-dependent Lamé-parameters is already an ill-posed problem, this problem should also be ill-posed. Similar to the acoustic setting, we can improve the known results by showing that purely time-dependent Lamé-parameters lead to ill-posed problems as well.

We would like to emphasize that this analysis is motivated by mathematical curiosity. We are not certain if time-dependent parameters λ, μ are physical and whether the linear elastic equation would still be appropriate for this case. Furthermore, we only scratch the surface of elastic wave propagation. For example, we do not take into account that the elastic equation allows for different wave modes (transversal and longitudinal) propagating at different speeds that depend on the Lamé-parameters. For more details we refer the reader to the pertinent literature on continuum mechanics, e.g. [Cia88; TM05; Zei88].

ABSTRACT FORMULATION We start by stating the elastic equation in the required abstract setting. As motivated in the introduction, we consider the initial boundary value problem

$$(\rho u')' - \operatorname{div}(\sigma_{\lambda,\mu} u) = f \quad \text{in } (0,T) \times \Omega, \tag{6.2a}$$

$$u(0) = u'(0) = 0 \quad \text{in } \Omega, \tag{6.2b}$$

$$u = 0 \quad \text{on } (0,T) \times \partial\Omega. \tag{6.2c}$$

Due to the boundary conditions, the appropriate function spaces for the weak formulation are given through $H := L^2(\Omega; \mathbb{R}^3)$ and $V := H_0^1(\Omega; \mathbb{R}^3)$. Therefore, we have $V^* = H^{-1}(\Omega; \mathbb{R}^3)$. Using these spaces, the stress tensor σ can be regarded as

$$\sigma_{\lambda(t),\mu(t)} \in \mathcal{L}\big(H_0^1(\Omega; \mathbb{R}^3),\, L^2(\Omega;\, \mathbb{R}^{3\times 3})\big)$$
$$\sigma_{\lambda(t),\mu(t)}\varphi := 2\mu(t)\varepsilon(\varphi) + \lambda(t)\operatorname{div}(\varphi) I_3, \quad \varphi \in H_0^1(\Omega; \mathbb{R}^3).$$

A formal integration by parts shows that

$$\int_\Omega \operatorname{div}\bigl(\sigma_{\lambda(t),\mu(t)}\varphi\bigr)\cdot\psi\,dx = -\int_\Omega \sigma_{\lambda(t),\mu(t)}\varphi(t) : \varepsilon(\psi)\,dx \tag{6.3}$$

holds for $\psi \in H^1_0(\Omega;\mathbb{R}^3)$. The expression $U : V$ refers to the scalar product of the matrices $U, V \in \mathbb{R}^{3\times 3}$, i.e. $U : V = \sum_{i,j} U_{ij}V_{ij}$. Naturally, the left-hand side of (6.3) is only well-defined in the classical sense for twice differentiable φ and once differentiable $\lambda(t)$, $\mu(t)$. However, we can use the right-hand side of (6.3) to regard the row-wise divergence operator in a distributional sense, that is $\operatorname{div}\colon L^2\bigl(\Omega;\mathbb{R}^{3\times 3}\bigr) \to H^{-1}(\Omega;\mathbb{R}^3)$.

In consequence, the weak formulation of the classical problem (6.2) reads

$$\int_\Omega \frac{d}{dt}\bigl(\rho(t)u'(t)\bigr)\cdot\psi\,dx + \int_\Omega \sigma_{\lambda(t),\mu(t)}u(t) : \varepsilon(\psi)\,dx = \int_\Omega f(t)\cdot\psi\,dx,$$

which should hold for all $\psi \in H^1_0(\Omega;\mathbb{R}^3)$ and almost all $t \in I$. To write this formulation in the abstract setting, we define $A_{\lambda,\mu}$ and C_ρ through

$$\begin{aligned}\bigl\langle A_{\lambda,\mu}(t)\psi,\varphi\bigr\rangle &\coloneqq \int_\Omega \sigma_{\lambda(t),\mu(t)}\psi : \varepsilon(\varphi)\,dx\\ C_\rho(t)v &\coloneqq \rho(t)v\end{aligned}$$

for all $v \in L^2(\Omega;\mathbb{R}^3)$, $\psi,\varphi \in H^1_0(\Omega;\mathbb{R}^3)$ and almost all $t \in I$. It is obvious that the operator $C_\rho(t)$ is self-adjoint. Further, we see that

$$\bigl\langle A_{\lambda,\mu}(t)\psi,\varphi\bigr\rangle = \int_\Omega 2\mu(t)\varepsilon(\psi) : \varepsilon(\varphi) + \lambda(t)\operatorname{div}(\psi)\operatorname{div}(\varphi)\,dx,$$

which also shows this for A.

To conclude, we are interested in the function $u \in L^2(I;H^1_0(\Omega;\mathbb{R}^3)) \cap H^1(I;L^2(\Omega;\mathbb{R}^3))$ with $\mathcal{C}u' \in H^1(I;H^{-1}(\Omega;\mathbb{R}^3))$ which solves

$$\begin{aligned}(\mathcal{C}_\rho u')' + \mathcal{A}_{\lambda,\mu}u = f \text{ in } L^2(I;H^{-1}(\Omega;\mathbb{R}^3)),\\ u(0) = 0 \text{ in } L^2(\Omega;\mathbb{R}^3) \text{ and } (\mathcal{C}_\rho u')(0) = 0 \text{ in } H^{-1}(\Omega;\mathbb{R}^3).\end{aligned}$$

This problem fits into the abstract framework. Note that we have implicitly set the operators B and Q to zero because we do not need to make use of them. Thus, through Chapter 4 we gain results for the operator S that maps the tuple $(A_{\lambda,\mu}, C_\rho)$ to u. In particular, they let us conclude that $S\colon D(S) \subset X \to Y$ is well-defined for $f \in L^2(I;L^2(\Omega;\mathbb{R}^3))$ or $f \in H^1(I;H^{-1}(\Omega;\mathbb{R}^3))$. The Banach spaces X and Y are equal to

$$\begin{aligned}X &\coloneqq W^{1,\infty}\bigl(I;\mathcal{L}^{sa}\bigl(H^1_0(\Omega;\mathbb{R}^3), H^{-1}(\Omega;\mathbb{R}^3)\bigr)\bigr)\\ &\qquad\times W^{1,\infty}\bigl(I;\mathcal{L}^{sa}\bigl(L^2(\Omega;\mathbb{R}^3)\bigr)\bigr),\\ Y &\coloneqq L^\infty\bigl(I;H^1_0(\Omega;\mathbb{R}^3)\bigr) \cap W^{1,\infty}\bigl(I;L^2(\Omega;\mathbb{R}^3)\bigr)\end{aligned}$$

and the open domain of definition of S is given as

$$\begin{aligned}D(S) \coloneqq \Bigl\{ (A,C) \in X \Bigm| & A \in L^\infty(I;\mathcal{L}^{sa}_{a_0+\varepsilon}(H^1_0(\Omega;\mathbb{R}^3), H^{-1}(\Omega;\mathbb{R}^3))) \text{ and}\\ & C \in L^\infty(I;\mathcal{L}^{sa}_{c_0+\varepsilon}(L^2(\Omega;\mathbb{R}^3))) \text{ for some } \varepsilon > 0 \Bigr\}.\end{aligned}$$

For a smooth $f \in \mathcal{F}^{(k)}$ (as defined on page 48) we can also regard S for $k \in \mathbb{N}_0$ as $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$, where

$$\begin{aligned} X^{(k)} &:= W^{k+1,\infty}\big(I; \mathcal{L}^{sa}\big(H_0^1(\Omega;\mathbb{R}^3),\, H^{-1}(\Omega;\mathbb{R}^3)\big)\big) \\ &\qquad \times W^{k+1,\infty}\big(I; \mathcal{L}^{sa}\big(L^2(\Omega;\mathbb{R}^3)\big)\big), \\ Y^{(k)} &:= W^{k,\infty}\big(I; H_0^1(\Omega;\mathbb{R}^3)\big) \cap W^{k+1,\infty}\big(I; L^2(\Omega;\mathbb{R}^3)\big). \end{aligned}$$

We would like to compose S with the operator

$$P(\lambda,\mu,\rho) := (A_{\lambda,\mu}, C_\rho)$$

to obtain the forward operator $F = S \circ P$ of our problem. It is well-defined for those tuples (λ,μ,ρ) that are mapped onto $D(S) \cap X^{(k)}$ by P. As in the acoustic case, it is easy to see that if all of these parameters belong to $W^{k+1,\infty}(I; L^\infty(\Omega))$, then $P(\lambda,\mu,\rho)$ will be an element of $X^{(k)}$. Through the following Lemma we aim to understand which constraints on the parameters are needed to ensure $P(\lambda,\mu,\rho) \in D(S)$.

Lemma 6.1. *Let $k \geqslant 0$ and λ, μ, $\rho \in W^{k+1,\infty}(I; L^\infty(\Omega))$. We further assume*

$$\rho(t,x) \geqslant \rho_0,\ \ \mu(t,x) \leqslant \alpha_0 \text{ and } \alpha_0^{-1} \leqslant 2\mu(t,x) + 3\lambda(t,x) \leqslant \alpha_0$$

for almost all $(t,x) \in I \times \Omega$ for some ρ_0, $\alpha_0 > 0$. Then

$$\big\langle A_{\lambda,\mu}(t)\varphi, \varphi\big\rangle \geqslant \alpha_0 \|\varepsilon(\varphi)\|^2_{L^2(\Omega;\mathbb{R}^{3\times3})} \ \text{ and } \ (C_\rho(t)\nu,\nu) \geqslant \rho_0\|\nu\|^2_{L^2(\Omega;\mathbb{R}^3)}$$

holds for all $\varphi \in H_0^1(\Omega;\mathbb{R}^3)$, $\nu \in L^2(\Omega;\mathbb{R}^3)$ and almost all $t \in I$.

Proof. Showing the coercivity of $C(t)$ is straightforward. It is obvious that $A(t)$ is coercive if both Lamé-parameters are bounded away from zero. However, the calculations in [KR16] prove that the conditions given in the assertion are already sufficient. □

Let $\rho_0 > 0$ and $\alpha_0 > 0$ be fixed in the sequel. The previous lemma only yields coercivity of A with respect to the L^2-norm of symmetric strain tensor $\varepsilon(\varphi) = (Du + Du^\top)/2$. Using Korn's inequality (found as Theorem 10.1 in [McL00]) we obtain a constant $C_K > 0$ such that $\|\varepsilon(\varphi)\|^2_{L^2(\Omega;\mathbb{R}^{3\times3})} \geqslant C_K\|D\varphi\|^2_{L^2(\Omega;\mathbb{R}^{3\times3})}$. Finally, this is connected to the $H_0^1(\Omega;\mathbb{R}^3)$-norm of φ by $\|D\varphi\|^2_{L^2(\Omega;\mathbb{R}^{3\times3})} \geqslant C_P\|\varphi\|^2_{H_0^1(\Omega;\mathbb{R}^3)}$, where the positive constant C_P is provided by Poincaré's inequality. Thus,

$$P\colon D(P) \cap W^{(k)} \to D(S) \cap X^{(k)}$$

is well-defined for $k \geqslant 0$ if we set the constants that appear in the definition of $D(S)$ to be $c_0 := \rho_0$, $a_0 := \alpha_0 C_K C_P$ and further define

$$\begin{aligned} W^{(k)} &:= \big(W^{k+1,\infty}(I; L^\infty(\Omega))\big)^3, \\ D(P) &:= \Big\{ (\lambda,\mu,\rho) \in W^{(0)} \ \Big|\ \alpha_0^{-1} + \varepsilon \leqslant 2\mu(t,x) + 3\lambda(t,x) \leqslant \alpha_0 - \varepsilon, \\ &\qquad\qquad \rho(t,x) \geqslant \rho_0 + \varepsilon \text{ and } \mu(t,x) \leqslant \alpha_0 - \varepsilon \\ &\qquad\qquad \text{a.e. in } I \times \Omega \text{ for some } \varepsilon > 0 \Big\}. \end{aligned}$$

The forward operator can therefore again be considered as

$$F := S \circ P\colon D(P) \cap W^{(k)} \to Y^{(k)}$$

for arbitrary $k \in \mathbb{N}_0$.

6.1.1 *Differentiability*

We already know about the differentiability of S. In our elastic setting P is linear. For $x = (\lambda, \mu, \rho) \in W^{(k)}$ we see that the norm of $(A_{\lambda,\mu}, C_\rho)$ in $X^{(k)}$ is bounded by the norm of x, since

$$\begin{aligned} \|A_{\lambda,\rho}\| &\leqslant 2\|\mu\|_{W^{k+1,\infty}(I;L^\infty)} + \|\lambda\|_{W^{k+1,\infty}(I;L^\infty)}, \\ \|C_\rho\| &\leqslant \|\rho\|_{W^{k+1,\infty}(I;L^\infty)}. \end{aligned}$$

This proves that P is bounded and therefore differentiable with constant derivative $\partial P(x)[h] = P(h)$ for $h \in W^{(k)}$. By using this we can obtain the following differentiability result for F.

Theorem 6.2. *Let* $k \geqslant 2$ *and* $f \in \mathcal{F}^{(k)}$. *Then* $F\colon D(P) \cap W^{(k)} \to Y^{(k-2)}$ *is Fréchet-differentiable. For all* $x = (\lambda, \mu, \rho) \in D(P) \cap W^{(k)}$ *and* $h = (\bar\lambda, \bar\mu, \bar\rho) \in W^{(k)}$, *the value* $\partial F(x)[h]$ *is given as the unique weak solution* $u_h \in Y^{(k-1)}$ *of*

$$(\rho u_h')'(t) - \operatorname{div}\big(\sigma_{\lambda(t),\mu(t)} u_h(t)\big) = \operatorname{div}\Big(\sigma_{\bar\lambda(t),\bar\mu(t)} u(t)\Big) - (\bar\rho u')'(t) \tag{6.4}$$

in $L^2(I; H^{-1}(\Omega;\mathbb{R}^3))$, *together with homogeneous initial values* $u_h(0) = (\rho u_h')(0) = 0$. *As always,* $u = F(x)$ *denotes the solution of the forward problem.*

Proof. The assertion follows by forming $\partial S(P(x))[P(h)]$ according to Corollary 4.8. The right-hand side is equal to

$$-A_{\bar\lambda,\bar\mu}(t)u(t) - (\mathcal{C}_{\bar\rho} u')'(t) \in H^{-1}(\Omega;\mathbb{R}^3),$$

when the second-order spatial differential operators in $A_{\bar\lambda,\bar\mu}(t)$ are understood in the variational sense using (6.3). □

We continue with the adjoint of $\partial F(x) \in \mathcal{L}(W^{(k)}, L^2(I; L^2(\Omega;\mathbb{R}^3)))$. We already know about $\partial S(P(x))^*$, and due to the chain rule (and linearity of P) we have

$$\partial F(x)^* = P^* \circ \partial S(P(x))^*.$$

Again we identify the dual space of $L^2(I; L^2(\Omega;\mathbb{R}^3))$ with the space itself. With a similar approach as used for Theorem 5.5, we obtain the following result.

Theorem 6.3. *Let* $k \geqslant 2$, $f \in \mathcal{F}^{(k)}$ *and* $x \in D(P) \cap W^{(k)}$. *The application of the adjoint of* $\partial F(x) \in \mathcal{L}\big(W^{(k)}, L^2(I; L^2(\Omega;\mathbb{R}^3))\big)$ *on* $v \in L^2(I; L^2(\Omega;\mathbb{R}^3))$ *can be written as*

$$\partial F(x)^*[v] = \begin{pmatrix} -\operatorname{div} u \operatorname{div} w_v \\ -2\varepsilon(u) : \varepsilon(w_v) \\ u' \cdot w_v' \end{pmatrix} \in L^1(I; L^1(\Omega))^3,$$

where the embedding of $L^1(I; L^1(\Omega))$ *into* $L^\infty(I; L^\infty(\Omega))^*$ *is understood using the inner product of* $L^2(I; L^2(\Omega))$ *and* $u = F(x) \in Y^{(2)}$. *With* $w_v \in Y$ *we denote the solution of*

$$(\rho w_v')'(t) - \operatorname{div}\big(\sigma_{\lambda(t),\mu(t)} w_v(t)\big) = v(t) \text{ in } H^{-1}(\Omega;\mathbb{R}^3),$$

which must hold for almost all $t \in I$*, together with homogeneous end conditions* $(\rho w_v')(T) = w_v(T) = 0$.

Proof. Let $h = (\bar\lambda, \bar\mu, \bar\rho) \in W^{(k)}$. Since the abstract formulation of our elastic wave equation has $B = Q = 0$, the adjoint equation (4.14a) is simply the original equation that has to be solved backwards in time.

We employ the characterization of $\partial S(P(x))^*$ from Corollary 4.10 to see

$$\begin{aligned}
\langle \partial F(x)^*[v], h\rangle_{(W^{(k)})^* \times W^{(k)}} &= \langle \partial S(P(x))^*[v], P(h)\rangle_{(X^{(k)})^* \times X^{(k)}} \\
&= \int_0^T \big(C_{\bar\rho} u'(t), w_v'(t)\big) - \big\langle A_{\bar\lambda,\bar\mu}(t) u(t), w_v(t)\big\rangle \, dt \\
&= \int_0^T \big(\bar\rho(t) u'(t), w_v'(t)\big) - \big(\bar\lambda(t) \operatorname{div} u(t), \operatorname{div} w_v(t)\big) \\
&\qquad - \int_\Omega 2\bar\mu(t) \varepsilon(u(t)) : \varepsilon(w_v(t)) \, dx \, dt.
\end{aligned}$$

Since e.g. $\bar\rho(t) \in L^\infty(\Omega)$ and $u'(t) \cdot w_v'(t) \in L^1(\Omega)$, we can also write this in a way that the integrands are dual products with the linearization parameters, i.e.

$$\begin{aligned}
\langle \partial F(x)^*[v], h\rangle_{(W^{(k)})^* \times W^{(k)}} = \int_0^T & \big\langle u'(t) \cdot w_v'(t), \bar\rho(t)\big\rangle \\
&- \big\langle \operatorname{div} u(t) \operatorname{div} w_v(t), \bar\lambda(t)\big\rangle \\
&- \langle 2\varepsilon(u(t)) : \varepsilon(w_v(t)), \bar\mu(t)\rangle \, dt.
\end{aligned}$$

Here, $\langle \cdot, \cdot \rangle$ denotes the dual product between $L^1(\Omega) \subset L^\infty(\Omega)^*$ and $L^\infty(\Omega)$. □

6.1.2 *Ill-posedness*

As in the acoustic case, we will again explicitly construct sequences of parameters, and then make use of Theorem 4.11 to prove that their images under F have to converge. Moreover, we can even re-use the helper functions α_j from Lemma 5.6.

Theorem 6.4. *Let* $k \in \mathbb{N}$ *and* $f \in \mathcal{F}^{(k)}$*. Then the task of finding* λ*,* μ *or* ρ *such that* $F(\lambda, \mu, \rho) = y \in Y^{(k-1)}$ *is locally ill-posed in every* $(\lambda, \mu, \rho) \in D(P) \cap W^{(k)}$.

Proof. Let $p = (\lambda, \mu, \rho) \in D(P) \cap W^{(k)}$ and $(A, C) = P(p) \in X^{(k)}$. Since $D(P) \cap W^{(k)}$ is an open subset of $W^{(k)}$, there exists $\delta_0 > 0$ with $B(p, \delta_0) \subset D(P) \cap W^{(k)}$. Let $0 < \delta \leqslant \delta_0$ be fixed and $(\alpha_j)_{j \in \mathbb{N}}$ be the sequence from Lemma 5.6 with $r := k + 1$.

(i) Identification of ρ: This part of the proof is similar to the identification of the wave speed c in the acoustic wave equation. However, it is much simpler here because ρ enters the equation in a linear fashion and it is only involved in one of the operators. We set $\rho_j(t, x) = \rho(t, x) + \delta \alpha_j(t)/2$ and $p_j := (\lambda, \mu, \rho_j)$. In effect, $\rho_j \in B(\rho, \delta)$ but

$\rho_j \not\to \rho$, both with respect to the norm of $W^{k+1,\infty}(I; L^\infty(\Omega))$. Note that $P(p_j) = (A, C + R_j^C)$ with

$$R_j^C(t)\varphi = \frac{\delta\alpha_j(t)}{2}\,\varphi \in L^2(\Omega; \mathbb{R}^3)$$

for all $\varphi \in L^2(\Omega; \mathbb{R}^3)$ and almost all $t \in I$. The norm of R_j^C stays bounded for $j \to \infty$ due to continuity of P. Moreover, for every fixed $u \in H^{k+1}(I; L^2(\Omega; \mathbb{R}^3))$ we see

$$\begin{aligned}\left\|(\mathcal{R}_j^C u')'\right\|_{H^{k-1}(I;L^2(\Omega;\mathbb{R}^3))} &\leqslant \left\|R_j^C u'\right\|_{H^k(I;L^2(\Omega;\mathbb{R}^3))} \\ &= \frac{\delta}{2}\left\|\alpha_j(\cdot)\left\|u'(\cdot)\right\|_{L^2(\Omega;\mathbb{R}^3)}\right\|_{H^k(I)} \to 0\end{aligned}$$

as $j \to \infty$. Using Theorem 4.11, we conclude that $F(p_j) = S(A, C + R_j^C) \to S(A, C) = F(p)$ in $Y^{(k-1)}$ as $j \to \infty$.

(ii) Identification of λ or μ: Continuing in the same fashion, we set $\lambda_j(t,x) = \lambda(t,x) + \delta\alpha_j(t)/2$ and $p_j := (\lambda_j, \mu, \rho)$. Hence, $P(p_j) = (A + R_j^A, C)$ with

$$R_j^A(t)\varphi = -\frac{\delta}{2}\alpha_j(t)\,\mathrm{div}(\mathrm{div}\,\varphi I_3) \in H^{-1}(\Omega; \mathbb{R}^3)$$

for all $\varphi \in H_0^1(\Omega; \mathbb{R}^3)$ and almost all $t \in I$. Again, the norm of R_j^A remains bounded as $j \to \infty$. For $v \in H^k(I; H_0^1(\Omega; \mathbb{R}^3))$ we see that

$$\left\|\mathcal{R}_j^A v\right\|_{H^k(I;H^{-1}(\Omega;\mathbb{R}^3))} \leqslant \frac{\delta}{2}\left\|\alpha_j(\cdot)\left\|v(\cdot)\right\|_{H_0^1(\Omega;\mathbb{R}^3)}\right\|_{H^k(I)} \to 0$$

for $j \to \infty$. This enables us to apply Theorem 4.11 once again. For μ we can do the same with $R_j^A(t)\varphi := -\delta\alpha_j(t)\,\mathrm{div}\,\varepsilon(\varphi) \in H^{-1}(\Omega; \mathbb{R}^3)$. □

We can show the following result concerning the compactness of the derivative of F.

Corollary 6.5. *Let $k \geqslant 2$ and $f \in \mathcal{F}^{(k)}$. We consider $F\colon D(P) \cap W^{(k)} \to Z$ with $Z = W^{j,p}(I; H)$ or $Z = C^j(I; H)$ for $0 \leqslant j \leqslant k$ and $1 \leqslant p < \infty$. For every $x = (\lambda, \mu, \rho) \in D(P) \cap W^{(k)}$, the linearization $\partial F(x) \in \mathcal{L}(W^{(k)}, Z)$ is a compact operator.*

Proof. $\partial F(x) = \partial S(P(x)) \circ P$ with linear and continuous P and compact $\partial S(P(x))$ (cf. Lemma 4.16). □

Similar to the acoustic case, the partial derivatives of F have discontinuous generalized inverses.

Lemma 6.6. *Let $k \geqslant 2$, $f \in \mathcal{F}^{(k)}$ and $f \neq 0$. For all $x \in D(P) \cap W^{(k)}$ the ranges of*

$$\partial_\lambda F(x),\ \partial_\mu F(x) \ \text{ and } \ \partial_\rho F(x) \in \mathcal{L}\Big(W^{k+1,\infty}(I; L^\infty(\Omega)),\, Y^{(k-1)}\Big)$$

are of infinite dimension.

Proof. Choose $(\beta_i)_{i\in\mathbb{N}} \subset C^\infty(I)$ in such a way that their supports are nonempty, pairwise disjoint and such that $\beta_i u$ does not vanish everywhere for all $i \in \mathbb{N}$ (possible, because $u \in C(I; L^2(\Omega;\mathbb{R}^3)) \setminus \{0\}$). We show that there exists a linearly independent set of right-hand sides to the linearized elastic wave equation (6.4). When considering the identification of ρ this is the case: The choice $\bar{\rho}_i(t,x) = \beta_i(t)$ yields the right-hand side $-(\beta_i u')'$, and since $u' \neq 0$ the β_i can be modified in such a way that $\{-(\beta_i u')'\}_{i\in\mathbb{N}}$ is linearly independent. For more details we refer to the proof of Lemma 5.9.

If we use $\bar{\lambda}_i(t,x) := \bar{\mu}_i(t,x) := \beta_i(t)$, the right-hand side of the linearized PDE with respect to λ reads $\beta_i(t)\operatorname{div}(\operatorname{div} u(t) I_3)$, which (by the construction of the β_i) yields a set of linearly independent functions if and only if $\operatorname{div}(\operatorname{div} u(t) I_3) \neq 0$. For μ, this right-hand side is equal to $2\beta_i(t)\operatorname{div}(\varepsilon(u(t)))$. Either $\operatorname{div}(\operatorname{div} u(t) I_3) = 0$ or $2\operatorname{div}(\varepsilon(u(t))) = 0$ would imply $u(t) = 0$. This can be seen by testing these (distributional) equalities with $u(t)$ and employing the coercivity of $A_{1,0}$ and $A_{0,1}$. □

6.2 MAXWELL'S EQUATIONS

As the third possible application of the abstract theory we choose a simple model based on Maxwell's equations. To be more precise, we deal with a second-order equation for the electrical field E inside a bounded domain $\Omega \subset \mathbb{R}^3$, which reads

$$(\varepsilon E')' + \operatorname{curl}\big(\mu^{-1}\operatorname{curl} E\big) = f. \tag{6.5a}$$

The equation is furnished with the initial- and boundary conditions

$$E(0) = E'(0) = 0 \quad \text{and} \qquad E = 0 \text{ on } (0,T)\times\partial\Omega. \tag{6.5b}$$

We aim to analyze the identification of a time- and space-dependent electric permittivity ε and magnetic permeability μ. The treatment of this problem is similar to both the acoustic- and the elastic wave equations, therefore we give an even less detailed discussion of this problem than in the previous section.

A derivation of (6.5) based on Maxwell's equations can be found in [Mon03]. A related first-order system with time-independent parameters has been analyzed in [KR16]. For further mathematical models and inverse problems in electrodynamics, mainly based on time-harmonic solutions, we direct the reader to [CK13; KH15]. For a general, physical interpretation of (6.5a) we refer to [Fey15]. Note that this equation seems to be physical for a time-dependent parameters as well. For example, the article [Hay+16] deals with such a scenario in the context of photonic structures. There, the permittivity ε is assumed to be time-dependent because of fast changes in the underlying medium.

Appropriate function spaces for the system (6.5) are

$$V := \big\{\, E \in H^1_0(\Omega;\mathbb{R}^3) \;\big|\; \operatorname{div} E = 0 \,\big\} \quad \text{and} \quad H := L^2(\Omega;\mathbb{R}^3).$$

For a first-order Maxwell system one would typically use the space $H_0(\operatorname{curl}, \Omega)$. Since we need curl^2 to be coercive on V, we make the additional assumptions that $\operatorname{div} E = 0$ and that not only the tangential component, but also the normal component of E vanishes on $\partial\Omega$. For smooth or convex Ω, the set of $H_0(\operatorname{curl}, \Omega)$ functions that fulfill these restrictions coincides with V (cf. [Mon03]). We endow V with the $H^1(\Omega; \mathbb{R}^3)$-norm and for notational purposes continue to abbreviate it using V.

The weak formulation can again be written in the form $(\mathcal{C}u')' + \mathcal{A}u = f$ with $u = E$. The operators $C = C_\varepsilon$, $A = A_\mu$ are given as

$$C_\varepsilon(t)v = \varepsilon(t)v \quad \text{and} \quad \langle A_\mu(t)\psi, \varphi\rangle = \Big(\mu(t)^{-1} \operatorname{curl}\psi, \operatorname{curl}\varphi\Big)_{L^2(\Omega;\mathbb{R}^3)}$$

for $v \in L^2(\Omega; \mathbb{R}^3)$, $\varphi, \psi \in V$ and $t \in I$.

The operator S, that maps A_μ and C_ε onto E, is almost the same as the one that was used in the elastic setting, only the function spaces are different: For $k \in \mathbb{N}_0$ we look at $S\colon D(S) \cap X^{(k)} \to Y^{(k)}$, where

$$\begin{aligned}
X^{(k)} &\coloneqq W^{k+1,\infty}(I; \mathcal{L}^{sa}(V, V^*)) \times W^{k+1,\infty}\big(I; \mathcal{L}^{sa}\big(L^2(\Omega; \mathbb{R}^3)\big)\big),\\
D(S) &\coloneqq \Big\{ (A, C) \in X^{(0)} \;\Big|\; C \in L^\infty\big(I; \mathcal{L}^{sa}_{c_0+\delta}\big(L^2(\Omega; \mathbb{R}^3)\big)\big) \text{ and}\\
&\qquad\qquad A \in L^\infty\big(I; \mathcal{L}^{sa}_{a_0+\delta}(V, V^*)\big) \text{ for some } \delta > 0 \Big\},\\
Y^{(k)} &\coloneqq W^{k,\infty}(I; V) \cap W^{k+1,\infty}(I; L^2(\Omega; \mathbb{R}^3)).
\end{aligned}$$

Of course the right-hand side f has to be regular enough, i.e. has to belong to $\mathcal{F}^{(k)}$. The value operator for this setting reads

$$P(\varepsilon, \mu) \coloneqq (A_\mu, C_\varepsilon),$$

and by forming $F = S \circ P$ we obtain the forward operator of this problem.

COERCIVITY Of course we want the image of P to be a subset of $D(S)$. We immediately see that

$$\langle A_\mu(t)\varphi, \varphi\rangle \geqslant \mu_1 \|\operatorname{curl}\varphi\|^2_{L^2(\Omega;\mathbb{R}^3)} \quad \text{and} \quad (C_\varepsilon(t)v, v) \geqslant \varepsilon_0 \|v\|^2_{L^2(\Omega;\mathbb{R}^3)}$$

holds for all $\varphi \in V$, $v \in L^2(\Omega; \mathbb{R}^3)$ and almost all $t \in I$ if $\varepsilon \geqslant \varepsilon_0$ and $\mu \leqslant \mu_1$ almost everywhere. To prevent μ^{-1} from becoming unbounded, we pose the additional constraint $\mu(t, x) \geqslant \mu_0$. Let the constants μ_0, μ_1 and ε_0 be fixed in the sequel. This already takes care of the coercivity of C_ε. Since V contains those functions from $H^1_0(\Omega, \mathbb{R}^3)$ that are divergence free, $\|\operatorname{curl}\cdot\|_{L^2}$ is equivalent to the $H^1_0(\Omega)$-norm. Indeed, for smooth $\varphi \in C^\infty_c(\Omega; \mathbb{R}^3)$ with $\operatorname{div}\varphi = 0$ we observe that

$$\begin{aligned}
\int_\Omega |\operatorname{curl}\varphi|^2 \,dx &= \int_\Omega \varphi(\operatorname{curl}^2\varphi - \nabla(\operatorname{div}\varphi))\,dx\\
&= -\int_\Omega \Delta\varphi \,dx = \int_\Omega |\nabla\varphi|^2 \,dx.
\end{aligned}$$

By approximation, $\|\operatorname{curl}\varphi\|^2_{L^2} = \|\nabla\varphi\|^2_{L^2} \geqslant C_P\|\varphi\|^2_V$ holds for all $\varphi \in V$. With $C_P > 0$ we again denote the Poincaré-constant of Ω. According these considerations, P given as

$$P\colon D(P)\cap W^{(k)} \to D(S)\cap X^{(k)}$$

is well-defined for $k \in \mathbb{N}_0$. For this, we set the constants that appear in the definition of $D(S)$ to be $C_0 \coloneqq \varepsilon_0$ and $A_0 \coloneqq \mu_1 C_P$, and introduce the sets

$$\begin{aligned} W^{(k)} &\coloneqq \left(W^{k+1,\infty}(I; L^\infty(\Omega))\right)^2, \\ D(P) &\coloneqq \Big\{ (\varepsilon,\mu) \in W^{(0)} \;\Big|\; \mu_1 - \delta \geqslant \mu \geqslant \mu_0 + \delta,\ \varepsilon \geqslant \varepsilon_0 + \delta \\ &\qquad\qquad \text{a.e. in } I\times\Omega \text{ for some } \delta > 0 \Big\}. \end{aligned}$$

The forward operator F can then be considered as the mapping

$$F \coloneqq S \circ P\colon D(P)\cap W^{(k)} \to Y^{(k)}$$

for arbitrary $k \in \mathbb{N}_0$. We note that $D(P)\cap W^{(k)}$ is an open subset of the Banach space $W^{(k)}$, so analyzing the Fréchet-differentiability of F and P makes sense.

6.2.1 *Differentiability*

In contrast to elastic equation in the preceding section we now have to deal with a nonlinear P. Its Fréchet-derivative is the topic of the following Lemma.

Lemma 6.7. *For every $k \in \mathbb{N}_0$, the operator $P\colon D(P)\cap W^{(k)} \to X^{(k)}$ is Fréchet-differentiable. Its derivative $\partial P\colon D(P)\cap W^{(k)} \to \mathcal{L}(W^{(k)}, X^{(k)})$ is for all $(\varepsilon,\mu) \in D(P)\cap W^{(k)}$ and linearization parameters $(\bar\varepsilon,\bar\mu) \in W^{(k)}$ given by*

$$\partial P(\varepsilon,\mu)[\bar\varepsilon,\bar\mu] = t \mapsto \begin{pmatrix} \varphi \in V \mapsto -\operatorname{curl}\left(\frac{\bar\mu(t)}{\mu(t)^2}\operatorname{curl}\varphi\right) \in V^* \\ \nu \in L^2(\Omega;\mathbb{R}^3) \mapsto \bar\varepsilon(t)\nu \in L^2(\Omega;\mathbb{R}^3) \end{pmatrix}.$$

Proof. The right-hand side consists of $\partial A_\mu[\bar\mu]$ and $C_{\bar\varepsilon}$. The former can be calculated using Lemma 5.1. □

An application of the chain rule yields the following theorem for the derivative of F.

Theorem 6.8. *Let $k \geqslant 2$ and $f \in \mathcal{F}^{(k)}$. Then $F\colon D(P)\cap W^{(k)} \to Y^{(k-2)}$ is Fréchet-differentiable. For all $x = (\varepsilon,\mu) \in D(P)\cap W^{(k)}$ and $h = (\bar\varepsilon,\bar\mu) \in W^{(k)}$, we can characterize $\partial F(x)[h]$ as the unique weak solution E_h of the equation*

$$(\varepsilon E_h')'(t) + \operatorname{curl}\left(\mu(t)^{-1}\operatorname{curl} E_h(t)\right) = \operatorname{curl}\left(\frac{\bar\mu(t)}{\mu(t)^2}\operatorname{curl} E(t)\right) - (\bar\varepsilon E')'(t)$$

that also satisfies homogeneous initial values $E_h(0) = (\varepsilon E_h')(0) = 0$. The right-hand side involves the solution $E = F(x)$ of the forward problem.

Proof. Follows from Corollary 4.8. Note that the right-hand side is equal to $-\partial A_\mu[\bar{\mu}](t)E(t) - (\mathcal{C}_{\bar{\varepsilon}} E')'(t)$. $\square$

We continue with the adjoint of $\partial F(x) \in \mathcal{L}(W^{(k)}, L^2(I; L^2(\Omega; \mathbb{R}^3)))$. Its characterization is obtained in the same way as for the elastic case.

Theorem 6.9. *Let* $k \geqslant 2$, $f \in \mathcal{F}^{(k)}$ *and* $x \in D(P) \cap W^{(k)}$. *The application of the adjoint of* $\partial F(x) \in \mathcal{L}(W^{(k)}, L^2(I; L^2(\Omega; \mathbb{R}^3)))$ *on* $v \in L^2(I; L^2(\Omega; \mathbb{R}^3))$ *can be written as*

$$\partial F(x)^*[v] = \begin{pmatrix} -\mu^{-2} \operatorname{curl} E \cdot \operatorname{curl} w_v \\ E' \cdot w_v' \end{pmatrix} \in L^1(I; L^1(\Omega))^2,$$

where $E = F(x) \in Y^{(k)}$ *and* $w_v \in Y^{(0)}$ *solves*

$$(\varepsilon w_v')'(t) + \operatorname{curl}(\mu(t)^{-1} \operatorname{curl} w_v(t)) = v(t)$$

in V^* *for almost all* $t \in I$ *and possesses homogeneous end conditions* $(\varepsilon w_v')(T) = w_v(T) = 0$.

Proof. Let $h = (\bar{\varepsilon}, \bar{\mu}) \in W^{(k)}$. Through the characterization of $\partial S(P(x))^*$ that is provided by Corollary 4.10 we conclude

$$\begin{aligned}
\langle \partial F(x)^*[v], h \rangle_{(W^{(k)})^* \times W^{(k)}} &= \langle \partial S(P(x))^*[v], P(h) \rangle_{(X^{(k)})^* \times X^{(k)}} \\
&= \int_0^T \left(C_{\bar{\varepsilon}} E'(t), w_v'(t) \right) - \langle \partial A_\mu[\bar{\mu}](t) E(t), w_v(t) \rangle \, dt \\
&= \int_0^T \left(\bar{\varepsilon}(t) E'(t), w_v'(t) \right) + \Big(\bar{\mu}(t) \mu(t)^{-2} \operatorname{curl} E(t), \operatorname{curl} w_v(t) \Big) \, dt.
\end{aligned}$$

For example, we have $\bar{\varepsilon}(t) \in L^\infty(\Omega)$ and $E'(t) \cdot w_v'(t) \in L^1(\Omega)$, thus we can reshape this into

$$\begin{aligned}
\langle \partial F(x)^*[v], h \rangle_{(W^{(k)})^* \times W^{(k)}} = \int_0^T & \Big\langle \mu(t)^{-2} \operatorname{curl} E(t) \cdot \operatorname{curl} w_v(t), \bar{\mu}(t) \Big\rangle \\
&+ \Big\langle E'(t) \cdot w_v'(t), \bar{\varepsilon}(t) \Big\rangle \, dt.
\end{aligned}$$

Here, $\langle \cdot, \cdot \rangle$ denotes the dual product between $L^1(\Omega)$ and $L^\infty(\Omega)$. $\square$

6.2.2 *Ill-posedness*

Unsurprisingly, the identification of ε and μ yields ill-posed problems.

Theorem 6.10. *Let* $k \in \mathbb{N}$ *and* $f \in \mathcal{F}^{(k)}$. *The task of finding* ε *or* μ *such that* $F(\varepsilon, \mu) = y \in Y^{(k-1)}$ *is locally ill-posed in every* $(\varepsilon, \mu) \in D(P) \cap W^{(k)}$.

Proof. Let $p = (\varepsilon, \mu) \in D(P) \cap W^{(k)}$ and $(A, C) = P(p) \in X^{(k)}$. Since $D(P) \cap W^{(k)}$ is an open subset of $W^{(k)}$, there exists $\delta_0 > 0$ with $B(p, \delta_0) \subset D(P) \cap W^{(k)}$. Let $0 < \delta \leqslant \delta_0$ be fixed and $(\alpha_j)_{j \in \mathbb{N}}$ be the sequence from Lemma 5.6 with $r := k + 1$. The ill-posedness of the reconstruction of ε

can be done using the exact same proof as for ρ in the elastic case. Thus, we only have to tackle the ill-posedness with respect to μ. We define

$$\mu_j(t,x) \coloneqq \frac{1}{\mu(t,x)^{-1} + \delta\alpha_j(t)/2} \quad \text{and} \quad p_j \coloneqq (\varepsilon, \mu_j).$$

This yields $P(p_j) = (A + R_j^A, C)$, where

$$R_j^A(t)\varphi = \frac{\delta}{2}\alpha_j(t)\operatorname{curl}^2 \varphi \in V^*$$

for all $\varphi \in V$ and almost all $t \in I$. Again, the norm of R_j^A stays bounded for $j \to \infty$, and for $v \in H^k(I;V)$ we see that

$$\|\mathcal{R}_j^A v\|_{H^k(I;V^*)} \leqslant \frac{\delta}{2}\big\|\alpha_j(\cdot)\|v(\cdot)\|_V\big\|_{H^k(I)} \to 0 \quad \text{as } j \to \infty.$$

This allows us to apply Theorem 4.11. □

Regarding the linearizations of F, we can prove the following analogs of Corollary 6.5 and Lemma 6.6.

Corollary 6.11. *Let* $k \geqslant 2$ *and* $f \in \mathcal{F}^{(k)}$. *We consider* $F\colon D(P) \cap W^{(k)} \to Z$ *with* $Z = W^{j,p}(I;H)$ *or* $Z = C^j(I;H)$ *for* $0 \leqslant j \leqslant k$ *and* $1 \leqslant p < \infty$. *For every* $x = (\varepsilon, \mu) \in D(P) \cap W^{(k)}$, *the linearization* $\partial F(x) \in \mathcal{L}\big(W^{(k)}, Z\big)$ *is a compact operator.*

Proof. $\partial F(x) = \partial S(P(x)) \circ \partial P(x)$ and $\partial S(P(x))$ is compact. □

Lemma 6.12. *Let* $k \geqslant 2$, $f \in \mathcal{F}^{(k)}$ *and* $f \neq 0$. *For all* $x \in D(P) \cap W^{(k)}$, *the ranges of*

$$\partial_\varepsilon F(x) \;\; \textit{and} \;\; \partial_\mu F(x) \in \mathcal{L}\Big(W^{k+1,\infty}(I;L^\infty(\Omega)),\, Y^{(k-1)}\Big)$$

are of infinite dimension.

Proof. The treatment of ε is identical to the treatment of ρ in the elastic case. Let $(\beta_i)_{i\in\mathbb{N}}$ denote the sequence constructed in the proof of Lemma 6.6. Choosing $\bar{\mu}_i(t,x) \coloneqq \beta_i(t)$, the right-hand side of the linearized PDE with respect to μ reads $\beta_i(t)\operatorname{curl}(\mu(t)^{-2}\operatorname{curl} E(t))$, which (by the construction of the β_i) yields a set of linearly independent functions if and only if $\operatorname{curl}(\mu(t)^{-2}\operatorname{curl} E(t)) \neq 0$ for at least one $t \in \operatorname{spt}\beta_i$. The contrary would imply

$$0 = \int_\Omega \mu(t)^{-2}|\operatorname{curl} E(t)|^2 \,dx \geqslant \mu_1^{-2}\int_\Omega |\operatorname{curl} E(t)|^2 \,dx$$

which would then lead us to conclude $E(t) = 0$. This cannot be the case for those t where $\beta_i(t) \neq 0$, because the β_i were constructed in such a way that $E(t) \neq 0$ for all $t \in \operatorname{spt}\beta_i$. □

7 DISCRETIZATION OF THE ACOUSTIC WAVE EQUATION

In the preceding chapters we have collected a lot of theoretical results about dynamic inverse problems that concern evolution equations. However, we also wish to shed a light on how (and how well) these problems can be solved in practice.

We choose the acoustic wave equation from Chapter 5 for these numerical experiments. The main reason is the fact that its solution involves far less unknowns than the vector-valued wave fields we have encountered for both the elastic wave equation and the model for electromagnetic waves. Furthermore, our analysis of the acoustic wave equation is not limited to the time-consuming three-dimensional case. Instead, it allows us to perform the majority of experiments in 2D, while giving only a few 3D results that expose dimension-related effects.

Obviously we need to be able to represent the data, the unknowns and the involved operators in a discrete setting. This chapter is devoted to these tasks. To be more precise, we need to be able to evaluate the forward operator F and its derivative ∂F. Furthermore, solving linearized problems (that will arise in the inversion algorithm) also involves evaluating the adjoint ∂F^*. From Chapter 5 we know that each of these operators requires the solution of wave equations. The first section of this chapter deals with the setup of a solver for the acoustic wave equation. In doing so, we will see which discrete representation of the wave field $u = F(x)$ and the parameters $x = (c, \nu, \rho, q)$ are the most convenient from a numerical point of view. Section 7.2 is concerned with the calculation of norms and inner products on these discretized quantities. By combining the PDE solver with these discrete norms, we can obtain approximations of the needed operators, as we will present in Section 7.3. In particular, there we discuss two possible approaches for the efficient evaluation of the adjoint ∂F^*. This will conclude the discretization, and will set the stage for the next chapter, where we are going to use of this setup for the actual reconstruction of parameters.

7.1 SOLVER FOR THE PARTIAL DIFFERENTIAL EQUATION

The basis of the theoretical considerations in Chapter 5 is the initial- and boundary value problem (5.15), which reads

$$\left(\frac{1}{\rho c^2}u'\right)' + \left(\frac{\rho'}{\rho^2 c^2} + \nu\right)u' - \operatorname{div}\frac{\nabla u}{\rho} + qu = f \qquad \text{in } (0,T)\times\Omega, \tag{7.1a}$$

$$u(0) = \left(c^{-2}\rho^{-1}u'\right)(0) = 0 \qquad \text{in } \Omega, \tag{7.1b}$$

$$u = 0 \qquad \text{on } (0,T)\times\partial\Omega. \tag{7.1c}$$

From Theorems 5.4 and 5.5 we know that the solution of such a problem is not only needed to evaluate the forward operator $F\colon (c,\nu,\rho,q)\mapsto u$ but also for its Fréchet-derivatives and the corresponding adjoints. However, we should not rely on (7.1) for our numerical considerations, because it makes explicit use of the first time derivative of the parameter ρ. This means that if ρ is to be reconstructed, then we have to set up another numerical scheme to approximate ρ' (e.g. by finite differences), resulting in an additional source of discretization errors. Alternatively, we could try to reconstruct the pair (ρ,ρ'). Naturally, this huge increase in the number of unknowns would negatively impact the reconstruction quality.

We can avoid these difficulties by directly working with problem (5.14); it involves the PDE

$$\frac{1}{\rho}\left(\frac{1}{c^2}u'\right)' + \nu u' - \operatorname{div}\frac{\nabla u}{\rho} + qu = f.$$

This is also the form in which we derived the acoustic wave equation. The weak formulation of this equation can be written as

$$\mathcal{D}^{-1}(\mathcal{C}u')' + \mathcal{B}u' + \mathcal{A}u = f \;\text{ in } L^2(I;H^{-1}(\Omega)), \tag{7.2}$$

with $D(t)v := \rho(t)v$, $C(t)v := v/c(t)^2$ and $B(t)v := \nu(t)v$ for $v\in L^2(\Omega)$. Separating $qu - \operatorname{div}\frac{\nabla u}{\rho}$ into two operators is only advantageous for the theoretical analysis, thus we set

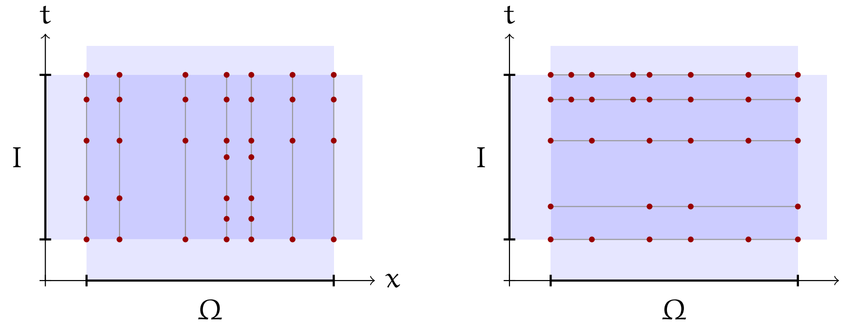
$$\langle A(t)\varphi,\psi\rangle := \int_\Omega \frac{\nabla\varphi\cdot\nabla\psi}{\rho(t)} + q(t)\varphi\psi\,dx, \quad \varphi,\psi\in H_0^1(\Omega).$$

Consequently, this $A(t)$ is (in general) only $H_0^1(\Omega)$-coercive up to a perturbation in $L^2(\Omega)$, i.e. it satisfies the Gårding-inequality

$$\langle A(t)\varphi,\varphi\rangle \geqslant A_0\|\varphi\|^2_{H_0^1(\Omega)} - \|q\|_{L^\infty(I;L^\infty)}\|\varphi\|^2_{L^2(\Omega)}.$$

Here, $A_0 = \rho_1^{-1}C_p^{-2}$ denotes the coercivity constant of the original A.

For conditions that ensure the unique solvability of initial value problems for (7.2), and its equivalence to (7.1) we can rely on Corollary 3.12. By using it we see that both equations are equivalent if $(c,\nu,\rho,q)\in W^{(1)}$, or in the case of a ρ that is constant in time.



(a) vertical lines (b) horizontal lines

Figure 7.1: Structuring of space-time meshes of line methods

In contrast to elliptic equations, hyperbolic and parabolic partial differential equations behave very differently in one dimension (often labeled as the time variable) than in all of the other (space) dimensions. This is not only a main theme in their theoretical analysis, but also in the numerical approximation of their solution. *Line methods* first perform a discretization in only one variable, yielding a semidiscrete problem, before also discretizing in the second variable, in the hope that this two-step process can be analyzed and discretized more efficiently. These methods naturally fall into two categories:

(i) *Vertical lines* (often just denoted as the "method of lines"): First discretize in space (for example using finite elements or finite differences). This results in a very large system of ODEs that can be solved with appropriate methods, for instance with Runge-Kutta schemes.

(ii) *Horizontal lines* (also known as the "Rothe-Method"): First perform a time discretization (e.g. by the Θ-method or the Newmark scheme). One obtains a sequence of elliptic boundary value problems that can then be discretized in space.

The names arise by picturing a x-t coordinate system, as shown in Figure 7.1. Discretization in time first yields a continuous problem at finitely many points in time, i.e. the space-time mesh is structured horizontally. Analogously, the other method corresponds to a series of vertical lines. For details we refer to the standard literature, like [GR05].

It is also possible to forgo this process and discretize the whole problem directly, for instance using finite elements in space-time. Typically, the memory needed to store and solve the resulting linear systems is much higher when compared to line methods. However, the price decline of computer memory has fueled their popularity, in the hope of achieving faster PDE solutions (the classical memory versus computing power trade-off). We want to be able to run multiple PDE solutions in parallel on a workplace computer (with 32 GiB of memory) and therefore decide to rely on a line method.

In literature, the method of vertical lines is featured more prominently (e.g. [DL00]). If the meshes are kept constant in time, then both line methods yield the same discretization (cf. [BGR10; BR99]). The Rothe-Method (horizontal lines) allows the use of time-dependent meshes, as long as one is able to transfer a function of one mesh to the next. Moreover, horizontal lines are a better fit with our theoretical analysis based on evolution equations: The semidiscretization in time does not rely on specific knowledge about spaces $V = H_0^1(\Omega)$ and $H = L^2(\Omega)$. Thus, at least the first part of the PDE solver is easily adapted to other equations, like the elastic wave equation. Therefore we will perform the discretization in the Rothe-framework.

7.1.1 *Semidiscretization in time*

Here, it is more convenient to work with (7.2) as a system of two first-order equations. By writing $v := u'$ we obtain the system

$$u' = v, \tag{7.3a}$$

$$\mathcal{D}^{-1}(\mathcal{C}v)' = (f - \mathcal{A}u - \mathcal{B}v), \tag{7.3b}$$

where the equalities should hold in $L^2(I; L^2(\Omega))$. There is nothing to be gained in the numerical analysis by requiring homogeneous initial values, therefore we complement this system by initial conditions

$$u(0) = u_0 \in H_0^1(\Omega) \quad \text{and} \quad v(0) = v_0 \in L^2(\Omega). \tag{7.3c}$$

We assume that there exists a pair (u, v) consisting of

$$u \in L^2(I; H_0^1(\Omega)) \cap H^1(I; L^2(\Omega)) \quad \text{and} \quad v \in H^1(I; L^2(\Omega))$$

that solves (7.3), and now aim to approximate this tuple.

Note that although $f(t)$ and $-A(t)u(t)$ may only belong to $H^{-1}(\Omega)$, at least their sum belongs to $L^2(\Omega)$, because the left-hand side of (7.3b) is an element of $L^2(\Omega)$. We make the equation explicit in $\mathcal{C}v$ by applying $\mathcal{D}$ to both sides of (7.3b).

Let $N \in \mathbb{N}$ and $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$ be a partition of $I = [0, T]$. At least formally, we may set $u^n := u(t_n)$ and $v^n := v(t_n)$ and in the same way introduce A^n, B^n, C^n, D^n and f^n for $n = 0, \ldots, N$. For the time derivatives of u and v we use the θ-scheme, which means that they are replaced by a weighted sum of forward- and backward difference quotients

$$\begin{aligned}
\frac{u^{n+1} - u^n}{t_{n+1} - t_n} &\approx \theta\, u'(t_{n+1}) + (1-\theta)\, u'(t_n), \\
\frac{C^{n+1}v^{n+1} - C^n v^n}{t_{n+1} - t_n} &\approx \theta\, (\mathcal{C}v)'(t_{n+1}) + (1-\theta)\, (\mathcal{C}v)'(t_n).
\end{aligned}$$

By inserting equations (7.3a) and (7.3b) into these expressions, we obtain the semidiscrete system

$$\frac{u^{n+1}-u^n}{t_{n+1}-t_n} = \theta v^{n+1} + (1-\theta)v^n, \tag{7.4a}$$

$$\begin{aligned}\frac{C^{n+1}v^{n+1}-C^n v^n}{t_{n+1}-t_n} &= \theta D^{n+1}\left(f^{n+1}-A^{n+1}u^{n+1}-B^{n+1}v^{n+1}\right)\\ &\quad + (1-\theta)D^n(f^n - A^n u^n - B^n v^n),\end{aligned} \tag{7.4b}$$

which we aim to solve for all $n = 0, \dots, N-1$ to obtain u^{n+1} and v^{n+1} based on the known u^n, v^n. The parameter $\theta \in [0,1]$ allows to control how implicit this scheme is with respect to u^{n+1} and v^{n+1}. If $\theta = 0$, then the right-hand sides do not depend on these values at all.

Some reorganization, and the definition of $k_{n+1} := (t_{n+1}-t_n)^{-1}$ shows (7.4) to be equivalent to

$$\begin{aligned}k_{n+1}u^{n+1} &= (1-\theta)v^n + k_{n+1}u^n + \theta v^{n+1}\\ \left(k_{n+1}C^{n+1}+\theta D^{n+1}B^{n+1}\right)v^{n+1} &= \theta D^{n+1}\left(f^{n+1}-A^{n+1}u^{n+1}\right)\\ &\quad + (1-\theta)D^n(f^n - A^n u^n - B^n v^n)\\ &\quad + k_{n+1}C^n v^n.\end{aligned}$$

If we eliminate v^{n+1} from the first equation, then we can use it to first solve for u^{n+1}, and then obtain v^{n+1} from the second equation. In order to do so, we must apply $k_{n+1}C^{n+1}+\theta D^{n+1}B^{n+1}$ (invertible for large k_{n+1}) to the first equation, which then attains the form

$$\begin{aligned}&k_{n+1}\left(k_{n+1}C^{n+1}+\theta D^{n+1}B^{n+1}\right)u^{n+1}\\ &\quad = \left(k_{n+1}C^{n+1}+\theta D^{n+1}B^{n+1}\right)\left[(1-\theta)v^n + k_{n+1}u^n\right]\\ &\qquad + \theta^2 D^{n+1}\left(f^{n+1}-A^{n+1}u^{n+1}\right)\\ &\qquad + \theta(1-\theta)D^n\left(f^n - A^n u^n - B^n v^n\right) + \theta k_{n+1}C^n v^n.\end{aligned}$$

Now it gets tricky. We wish to solve this equation for u^{n+1}, hence $A^{n+1}u^{n+1}$ belongs to the left-hand side. Unfortunately, we cannot evaluate D^{n+1} at f^{n+1} and $A^{n+1}u^{n+1}$ separately. Thus, the only way to proceed is applying $\left(D^{n+1}\right)^{-1}$ to both sides of this equation, which then allows all expressions that involve u^{n+1} to be moved to the left-hand side. This results in the system

$$\begin{aligned}&\left(k_{n+1}^2\left(D^{n+1}\right)^{-1}C^{n+1}+\theta k_{n+1}B^{n+1}+\theta^2 A^{n+1}\right)u^{n+1}\\ &\quad = \left(k_{n+1}\left(D^{n+1}\right)^{-1}C^{n+1}+\theta B^{n+1}\right)\left[(1-\theta)v^n + k_{n+1}u^n\right]\\ &\qquad + \theta\left(D^{n+1}\right)^{-1}\left[(1-\theta)D^n(f^n - A^n u^n - B^n v^n) + k_{n+1}C^n v^n\right]\\ &\qquad + \theta^2 f^{n+1},\end{aligned} \tag{7.5a}$$

$$\begin{aligned}&\left(k_{n+1}C^{n+1}+\theta D^{n+1}B^{n+1}\right)v^{n+1}\\ &\quad = \theta D^{n+1}\left(f^{n+1}-A^{n+1}u^{n+1}\right)\\ &\qquad + (1-\theta)D^n\left(f^n - A^n u^n - B^n v^n\right) + k_{n+1}C^n v^n.\end{aligned} \tag{7.5b}$$

This motivates the following semidiscrete algorithm for the solution of (7.2).

Algorithm 7.1 (Semidiscretization of (7.2))**.**

- Set $u^0 = u_0$ and $v^0 = v_0$.
- For every $n = 0, \dots, N-1$, calculate
  (1) $u^{n+1} \in H_0^1(\Omega)$ from u^n and v^n by solving (7.5a), and
  (2) $v^{n+1} \in L^2(\Omega)$ from u^{n+1}, u^n and v^n by solving (7.5b).

The first equation (7.5a) is a kind of Helmholtz-equation, because the operator on its left-hand side is the sum of an $H_0^1(\Omega)$-coercive operator (found inside A^{n+1}) and multiple operators belonging to $\mathcal{L}(L^2(\Omega))$. However, for large enough k_{n+1} the $L^2(\Omega)$-coercivity of $k_{n+1}C^{n+1}$ will force the whole operator to be $H_0^1(\Omega)$-coercive. By Lax-Milgram's lemma, this equation is then uniquely solvable for every right-hand side belonging to $H^{-1}(\Omega)$, and its solution u^{n+1} is an element of $H_0^1(\Omega)$. The operator in the second equation (7.5b) is $L^2(\Omega)$-coercive, but at first glance its right-hand side only belongs to $H^{-1}(\Omega)$, which would pose a problem. We sum up this discussion and take a closer look at the second equation in the following lemma.

Lemma 7.2. *Let $(c, \nu, \rho, q) \in D(F)$ with $\nu,\ q \in C(I; L^\infty(\Omega))$ and $f \in C(I; H^{-1}(\Omega))$. Further, let $\theta \in (0, 1]$, $u_0 \in H_0^1(\Omega)$, $v_0 \in L^2(\Omega)$ such that $f(0) - A(0)u_0 \in L^2(\Omega)$. We assume that the time step size $t_{n+1} - t_n = k_{n+1}^{-1}$ is small enough so that*

$$k_{n+1} \geqslant \max\left\{ c_1^2\Big(\alpha_0 + \theta\rho_1\|\nu\|_{L^\infty(I;L^\infty)}\Big),\ \frac{\theta^2\rho_1\|q\|_{L^\infty(I;L^\infty)}}{\alpha_0} \right\} \tag{7.6}$$

is satisfied for every $n = 0, \dots, N-1$ and a fixed $\alpha_0 > 0$. Then Algorithm 7.1 generates a well-defined finite sequence $((u^n, v^n))_{n=0,\dots,N} \subset H_0^1(\Omega) \times L^2(\Omega)$, and for every $n = 0, \dots, N-1$, the tuple $(u^n, v^n, u^{n+1}, v^{n+1})$ solves the system (7.4).

Proof. Before we begin we would like to remind the reader that the constants ρ_0, ρ_1 and c_1 bound the parameters ρ and c from below and above, and originate from the definition of $D(F)$ on page 78.

The restriction on the time step size causes $k_{n+1}C^{n+1} + \theta D^{n+1}B^{n+1}$ to be $L^2(\Omega)$-coercive, because

$$\begin{aligned}&\big((k_{n+1}C^{n+1} + \theta D^{n+1}B^{n+1})\varphi, \varphi\big)\\ &\qquad \geqslant \Big(k_{n+1}c_1^{-2} - \theta\rho_1\|\nu\|_{L^\infty(I;L^\infty)}\Big)\|\varphi\|_{L^2(\Omega)}^2 \geqslant \alpha_0\|\varphi\|_{L^2(\Omega)}^2\end{aligned}$$

holds for all $\varphi \in L^2(\Omega)$. Moreover, we can easily show that the operator involved in the first equation is $H_0^1(\Omega)$-coercive; for $\psi \in H_0^1(\Omega)$ we have

$$\begin{aligned}&\Big\langle\Big(k_{n+1}^2\big(D^{n+1}\big)^{-1}C^{n+1} + \theta k_{n+1}B^{n+1} + \theta^2 A^{n+1}\Big)\psi, \psi\Big\rangle\\ &\qquad \geqslant \theta^2 A_0\|\psi\|_{H_0^1(\Omega)}^2 + \Big(k_{n+1}\rho_1^{-1}\alpha_0 - \theta^2\|q\|_{L^\infty(I;L^\infty)}\Big)\|\psi\|_{L^2(\Omega)}^2\\ &\qquad \geqslant \theta^2 A_0\|\psi\|_{H_0^1(\Omega)}^2.\end{aligned}$$

We give an inductive proof that shows u^{n+1} and v^{n+1} to be well-defined. From the assumptions in the assertion we have $(u^0, v^0) \in H_0^1(\Omega) \times L^2(\Omega)$ and $f^0 - A^0 u^0 \in L^2(\Omega)$. We prove that from $(u^n, v^n) \in H_0^1(\Omega) \times L^2(\Omega)$ and $f^n - A^n u^n \in L^2(\Omega)$ follows that (7.5) possesses a unique solution $(u^{n+1}, v^{n+1}) \in H_0^1(\Omega) \times L^2(\Omega)$ that also has $f^{n+1} - A^{n+1} u^{n+1} \in L^2(\Omega)$.

We have already seen that the right-hand side of (7.5a) belongs to $H^{-1}(\Omega)$, and know that the operator on the left-hand side is $H_0^1(\Omega)$-coercive, which ensures the unique existence of $u^{n+1} \in H_0^1(\Omega)$ that solves this equation. By rearranging (7.5a) we observe that

$$\begin{aligned} -\theta^2\left(f^{n+1} - A^{n+1}u^{n+1}\right) \\ &= \left(k_{n+1}\left(D^{n+1}\right)^{-1} C^{n+1} + \theta B^{n+1}\right)\left[(1-\theta)v^n + k_{n+1}(u^n - u^{n+1})\right] \\ &\quad + \theta\left(D^{n+1}\right)^{-1}\left[(1-\theta)D^n(f^n - A^n u^n - B^n v^n) + k_{n+1}C^n v^n\right]. \end{aligned}$$

By assumption, $f^n - A^n u^n \in L^2(\Omega)$ and there are no further occurrences of A or f on the right-hand side. Thus, the left-hand side and due to $\theta \neq 0$ also $f^{n+1} - A^{n+1}u^{n+1}$ belong to $L^2(\Omega)$. We can use this to solve the second equation (7.5b), which as we now know has a right-hand side in $L^2(\Omega)$, and a $L^2(\Omega)$-coercive operator on the left. Hence, this equation admits a unique solution $v^{n+1} \in L^2(\Omega)$. By design, the system (7.5) is equivalent to (7.4). $\square$

The lower bound (7.6) for k_{n+1} depends on the arbitrary constant $\alpha_0 > 0$. The first expression is increasing in α_0, and the second is decreasing in α_0. Therefore, their maximum is minimal when both expressions are equal. By choosing α_0 in this way, we deduce that

$$k_{n+1} \geqslant \frac{\theta\|q\|}{\sqrt{1/4\|v\|^2 + c_1^{-2}\rho_1^{-1}\|q\|} - 1/2\|v\|}.$$

is the optimal bound that we can obtain from (7.6) in the case that v and q do not vanish everywhere.

Further, we would like to note that it would have been sufficient to consider the L^∞-norms of the negative parts $v^- := \min\{v, 0\}$ and $q^- := \min\{q, 0\}$ of v and q in (7.6). Thus, the bound for k_{n+1} could be drastically simplified to read $k_{n+1} \geqslant 0$ (which is satisfied anyway) if we assume both functions to be nonnegative.

It is almost obvious why we had to assume $\theta > 0$ in Lemma 7.2: For $\theta = 0$, the θ-scheme coincides with the explicit Euler method, which has the advantage of only requiring the solution of linear equations in $L^2(\Omega)$. However, this also implies that the algorithm cannot ensure $u^{n+1} \in H_0^1(\Omega)$.

For a detailed analysis of the θ-scheme we refer to the pertinent literature, e.g. [GR05] and also the articles [BGR10; BR99] that discuss a variety of possible ways to perform the time discretization (and connect them to Galerkin methods in space-time). The key takeaways from the theory are the fact that the θ-schemes are stable for $\theta \geqslant 1/2$, and that the method for $\theta = 1/2$ is of second order and is also known as the *Crank-Nicolson* method. At least in theory, for $\theta \neq 1/2$ the scheme is only of first order, but choosing $\theta = 1$ (implicit Euler method) can be used to reduce oscillations in the solution. Furthermore, $\theta = 2/3$ gets rid of the third order term in the convergence analysis, which might also result in better approximations in practice (where the constants in front of these terms actually matter).

7.1.2 *Discretization in space*

The semidiscrete problem (7.5) still relies on the solution of infinite-dimensional elliptic problems in $H_0^1(\Omega)$ and $L^2(\Omega)$. We will change this now by employing the finite element method (FEM). Note that it is not our goal to give a detailed introduction into finite elements. There is an abundance of books that do a good job on explaining the method and the inner workings of finite element libraries, e.g. [BR03; BS08; SS05]. We mainly apply the ideas of this method in order to replace the infinite-dimensional linear problems in Algorithm 7.1 by suitable matrices, and focus on how they can be solved efficiently.

We will begin with the assumption that the mesh does not change between time steps, but later in this section we are going to remark on what needs to be changed to facilitate time-dependent meshes. Let a finite subspace $V_h \subset V = H_0^1(\Omega)$ be given, with a basis $\varphi_1, \dots, \varphi_M$ where $M \in \mathbb{N}$. In practice, we use piecewise linear and globally continuous functions on a rectangular mesh approximation of Ω, which we assume to be a polygon to avoid boundary issues. In order to obtain sparse matrices, the basis functions should have small support. We will always use the nodal basis, which means each basis function has the value 1 on exactly one node and vanishes on the other nodes.

Our main challenge is the fact that matrix assembly is costly, often even slower than the solution of the resulting linear systems. The semidiscrete problem contains a lot of operators that need spatial discretization, and on top of that they even change in between time steps. In consequence, we have to reassemble most matrices after every time step. Caching matrices in between PDE solves is not an option due to memory concerns, and at least one of the matrices will change constantly in the inversion process. Hence, we should try to keep the number of different operator combinations that we need to discretize as low as possible. To this end, we apply $\left(D^{n+1}\right)^{-1}$ to (7.5b), and further introduce the auxiliary variables

$$\begin{aligned} x^n &:= (1-\theta)v^n + k_{n+1}u^n \\ y^n &:= (1-\theta)\left(D^{n+1}\right)^{-1} D^n (f^n - A^n u^n - B^n v^n) \\ &\qquad + k_{n+1}\left(D^{n+1}\right)^{-1} C^n v^n. \end{aligned}$$

With them, the system we need to solve in each step reads

$$\begin{aligned}\Big(k_{n+1}^2\big(D^{n+1}\big)^{-1}C^{n+1}+\theta k_{n+1}B^{n+1}+\theta^2A^{n+1}\Big)u^{n+1}\\ =\Big(k_{n+1}\big(D^{n+1}\big)^{-1}C^{n+1}+\theta B^{n+1}\Big)x^n+\theta^2f^{n+1}+\theta y^n\end{aligned}\tag{7.7a}$$

$$\Big(k_{n+1}\big(D^{n+1}\big)^{-1}C^{n+1}+\theta B^{n+1}\Big)v^{n+1}=\theta(f^{n+1}-A^{n+1}u^{n+1})+y^n,\tag{7.7b}$$

which we now only expect to hold in $V_h^*\supset H^{-1}(\Omega)$ for $n=0,\dots,N-1$. Similarly, we restrict our search for (u^k,v^k) $(k=0,\dots,N)$ to $V_h\times V_h$. Thus, there exist coefficients $U^k,V^k\in\mathbb{R}^M$ such that

$$u^k=\sum_{j=1}^M U_j^k\varphi_j\quad\text{and}\quad v^k=\sum_{j=1}^M V_j^k\varphi_j.$$

Additionally, we define a right-hand side vector $F^k:=\big(\langle f^k,\varphi_i\rangle\big)_i\in\mathbb{R}^M$, and matrices $\mathbb{A}^k,\mathbb{B}^k,\mathbb{C}^k\in\mathbb{R}^{M\times M}$ element-wise for $i,j=1,\dots,M$ via

$$\begin{aligned}\mathbb{A}_{ij}^k&:=\langle A^k\varphi_i,\varphi_j\rangle=\int_\Omega \operatorname{div}\Big(\frac{\nabla\varphi_i\cdot\nabla\varphi_j}{\rho(t_k)}\Big)+q(t_k)\varphi_i\varphi_j\,dx,\\ \mathbb{B}_{ij}^k&:=\big(B^k\varphi_i,\varphi_j\big)=\int_\Omega \nu(t_k)\varphi_i\varphi_j\,dx\text{ and}\\ \mathbb{C}_{ij}^k&:=\Big(\big(D^k\big)^{-1}C^k\varphi_i,\varphi_j\Big)=\int_\Omega\frac{\varphi_i\varphi_j}{\rho(t_k)c(t_k)^2}\,dx.\end{aligned}$$

Numerically, these values have to be computed using quadrature formulas. The auxiliary variable x^n is a linear combination of u^n and v^n, and therefore belongs to V_h. Its coefficients with respect to the finite element basis are $X^n:=(1-\theta)V^n+k_{n+1}U^n$. The numerical implementation of these vectors and matrices should be straightforward in most finite element libraries.

Due to linearity, the system for (u^{n+1},v^{n+1}) is solved if and only if their coefficients vectors satisfy

$$\begin{aligned}\Big[\big(k_{n+1}^2\mathbb{C}^{n+1}+\theta k_{n+1}\mathbb{B}^{n+1}+\theta^2\mathbb{A}^{n+1}\big)U^{n+1}\Big]_i\\ =\Big[\big(k_{n+1}\mathbb{C}^{n+1}+\theta\mathbb{B}^{n+1}\big)X^n\Big]_i+\theta^2F_i^{n+1}+\theta(y^n,\varphi_i),\end{aligned}\tag{7.8a}$$

$$\Big[\big(k_{n+1}\mathbb{C}^{n+1}+\theta\mathbb{B}^{n+1}\big)V^{n+1}\Big]_i=\theta\Big[F^{n+1}-\mathbb{A}^{n+1}U^{n+1}\Big]_i+(y^n,\varphi_i)\tag{7.8b}$$

for all $i=1,\dots,M$. What remains is to find a means to compute the inner products (y^n,φ_i). We have

$$\begin{aligned}(y^n,\varphi_i)=(1-\theta)\Big(\big(D^{n+1}\big)^{-1}D^n(f^n-A^nu^n-B^nv^n),\varphi_i\Big)\\ +k_{n+1}\Big(\big(D^{n+1}\big)^{-1}C^nv^n,\varphi_i\Big),\end{aligned}\tag{7.9}$$

and the second summand is equal to $k_{n+1}\left[\mathbb{C}^{n,n+1}V^n\right]_i$ if we define the additional matrix $\mathbb{C}^{n,n+1}$ with entries given for $i, j = 1, \ldots, M$ as

$$\mathbb{C}_{ij}^{n,n+1} := \left(\left(D^{n+1}\right)^{-1} C^n \varphi_i, \varphi_j\right) = \int_\Omega \frac{\varphi_i \varphi_j}{\rho(t_{n+1}) c(t_n)^2} \, dx.$$

In the special case $\left(D^{n+1}\right)^{-1} D^n = \mathrm{Id}$ (obtained by a time-independent ρ), the first summand on the right-hand side of (7.9) would be equal to $(1 - \theta)[F^n - \mathbb{A}^n U^n - \mathbb{B}^n V^n]_i$, which is easily implemented. Furthermore, in this situation $\mathbb{C}^{n,n+1} = \mathbb{C}^n$, hence assembly of the matrix $\mathbb{C}^{n,n+1}$ would become unnecessary. However, it is not apparent how to proceed in the general case: From our theoretical analysis of the semidiscretization, we know that $f^n - A^n u^n - B^n v^n \in L^2(\Omega)$, but it might not belong to V_h. We would need the latter to apply a matrix representation of $(D^{n+1})^{-1} D^n$ to it. Even if we were allowed to apply this operator to each summand of $f^n - A^n u^n - B^n v^n$ individually (e.g. if $f \in L^2(I; L^2(\Omega))$), this would force us to assemble two additional matrices and one additional vector in each time step.

To overcome the problems above, we choose to insert a projection onto V_h before the application of D^n in (7.9). Let $\alpha \in \mathbb{R}^M$ denote the coefficients of this projected value, i.e.

$$P_{V_h}(f^n - A^n u^n - B^n v^n) = \sum_{j=1}^{M} \alpha_j \varphi_j.$$

For $i = 1, \ldots, M$ these coefficients satisfy

$$\begin{aligned} \sum_{j=1}^{M} \alpha_j \left(\varphi_i, \varphi_j\right) &= (P_{V_h}(f^n - A^n u^n - B^n v^n), \varphi_i) \\ &= [F^n - \mathbb{A}^n U^n - \mathbb{B}^n V^n]_i, \end{aligned}$$

and the left-hand side is simply the ith component of $\mathbb{M}\alpha$, where $\mathbb{M}$ denotes the *mass matrix* with entries $\mathbb{M}_{ij} := \left(\varphi_i, \varphi_j\right)$. We conclude that we can obtain the coefficients α by solving the additional linear system $\mathbb{M}\alpha = F^n - \mathbb{A}^n U^n - \mathbb{B}^n V^n$. The subsequent application of $(D^{n+1})^{-1} D^n$ is realized through the matrix $\mathbb{D}^{n,n+1} \in \mathbb{R}^{M \times M}$, defined by

$$\mathbb{D}_{ij}^{n,n+1} := \left(\left(D^{n+1}\right)^{-1} D^n \varphi_i, \varphi_j\right) = \int_\Omega \frac{\rho(t_n)}{\rho(t_{n+1})} \varphi_i \varphi_j \, dx$$

for $i, j = 1, \ldots, M$. Inserting all of this into (7.9) yields

$$\begin{aligned} (y^n, \varphi_i) \approx Y_i^n := (1 - \theta)&\left[\mathbb{D}^{n,n+1} \mathbb{M}^{-1} (F^n - \mathbb{A}^n U^n - \mathbb{B}^n V^n)\right]_i \\ &+ k_{n+1}\left[\mathbb{C}^{n,n+1} V^n\right]_i, \end{aligned}$$

which we employ in (7.8) to finally arrive at the system

$$\begin{aligned} &(k_{n+1}^2 \mathbb{C}^{n+1} + \theta k_{n+1} \mathbb{B}^{n+1} + \theta^2 \mathbb{A}^{n+1}) U^{n+1} \\ &\qquad = (k_{n+1} \mathbb{C}^{n+1} + \theta \mathbb{B}^{n+1}) X^n + \theta^2 F^{n+1} + \theta Y^n, \end{aligned} \tag{7.10a}$$

$$(k_{n+1} \mathbb{C}^{n+1} + \theta \mathbb{B}^{n+1}) V^{n+1} = \theta F^{n+1} - \theta \mathbb{A}^{n+1} U^{n+1} + Y^n. \tag{7.10b}$$

It can be solved in every time step $n = 1, \dots, N$. The initial values (U^0, V^0) are the coefficients of the $L^2(\Omega)$-projection of (u_0, v_0) onto V_h, but it is more convenient to simply interpolate them at the mesh nodes. In practice, we will always deal with homogeneous initial values anyway, where this is of no consequence.

TIME-DEPENDENT MESHES The "correct" way of working with a different mesh for every time step, represented by finite-dimensional spaces $V_h^0, \dots, V_h^N \subset V$, would mean going back to (7.7). These equations should yield $u^{n+1}, v^{n+1} \in V_h^{n+1}$, but also depend on functions f^n, u^n, v^n on the previous finite element space V_h^n. Operations on the latter should be replaced by nonsquare matrices (with test functions in V_h^{n+1} and trial functions in V_h^n). Assembling these matrices is tedious because it is not immediately clear which basis functions from V_h^n and V_h^{n+1} couple (i.e. have intersecting supports). Even if this is possible, for instance because the meshes share the same hierarchy, it would mean that matrices for the same operator would have to be assembled multiple times but on different mesh combinations. This could quickly negate the positive effects of adaptive meshing.

It is much more practical to compute parts of the two right-hand sides, that depend on values at time t_n on the old mesh and then simply interpolate them on the new mesh. Afterwards, we may continue with the operations that depend on t_{n+1}. This approach has the striking advantage that at any given time, only one mesh has to be held in memory. A similar method (based on projections) is analyzed in [BS05]. Let $I^{n,n+1} \in \mathcal{L}(V_h^n, V_h^{n+1})$ denote the interpolation from the old- to the new finite element space, and M_n, M_{n+1} the dimensions of the two spaces. Hence, $I^{n,n+1}$ can be represented through a matrix $\mathbb{I}^{n,n+1} \in \mathbb{R}^{M_{n+1} \times M_n}$. Quantities on the old mesh are exclusively found in the vectors X^n and Y^n, thus instead of (7.10) we obtain the system

$$\begin{aligned}\left(k_{n+1}^2 \mathbb{C}^{n+1} + \theta k_{n+1} \mathbb{B}^{n+1} + \theta^2 \mathbb{A}^{n+1}\right) U^{n+1} \\ = \left(k_{n+1} \mathbb{C}^{n+1} + \theta \mathbb{B}^{n+1}\right) \mathbb{I}^{n,n+1} X^n + \theta^2 F^{n+1} + \theta \mathbb{I}^{n,n+1} Y^n\end{aligned} \tag{7.11a}$$

$$\left(k_{n+1} \mathbb{C}^{n+1} + \theta \mathbb{B}^{n+1}\right) V^{n+1} = \theta F^{n+1} - \theta \mathbb{A}^{n+1} U^{n+1} + \mathbb{I}^{n,n+1} Y^n. \tag{7.11b}$$

We sum up the steps needed to acquire a numerical approximation of the solution of (7.2) in the following algorithm.

Algorithm 7.3 (Solver for the Acoustic Wave Equation)**.**

- Interpolate (or project) the initial values u_0, v_0 into vectors U^0, V^0.
- Assemble the vector F^0 and the matrices $\mathbb{A}^0$, $\mathbb{B}^0$ and $\mathbb{C}^0$.
- For $n = 0, \dots, N-1$:
  (1) Assemble $\mathbb{C}^{n,n+1}$ and $\mathbb{D}^{n,n+1}$.
  (2) Calculate the vectors X^n and Y^n.
  (3) *(Optional)* Perform a mesh change and transfer X^n and Y^n to the new mesh. Clear all other matrices and vectors.

(4) Assemble the vector F^{n+1} and the matrices $\mathbb{A}^{n+1}$, $\mathbb{B}^{n+1}$ and $\mathbb{C}^{n+1}$.
(5) Calculate the right-hand side and system matrix for (7.11a).
(6) Solve (7.11a) to acquire U^{n+1}.
(7) Complete the right-hand side and system matrix for (7.11b).
(8) Solve (7.11b) to acquire V^{n+1}.

In this algorithm, we used the word "assemble" for tasks that involve numerical quadrature. All subsequent "calculations" then mainly consist of adding and scaling these precomputed vectors and matrices. Each time step involves the solution of up to three (one is hidden in Y^n) high-dimensional sparse linear systems, for which iterative solvers like GMRES and CG are well-suited. Since we only have to deal with symmetric matrices we use the latter, preconditioned with SSOR (symmetric successive over-relaxation, [SB05]).

The algorithm was formulated in a way that the memory usage is kept minimal: At every given time, we require memory for up to five vectors, six matrices and, most importantly, only one mesh. Due to the possible interactions between the finite elements we use, these matrices possess at most 9 structural nonzero entries per row when $d = 2$, which increases to 27 in a three-dimensional setting. Of course we will also have to store the solution vector $(U^0, \dots, U^N)$ for the inversion process. The fact that we will rarely go below $N = 128$ means that the PDE-solver has a memory footprint that is comparable to the footprint of such a space-time finite element vector. However, the latter is typically not accessed as often and could thus more easily be stored on slower media.

7.1.3 *Numerical validation*

We have implemented Algorithm 7.3 in C++, using the open-source finite-element library DEAL.II [Arn+19; BHK07]. It helps us to keep track of the mesh, manages the degrees of freedom on this mesh and performs numerical quadrature. The code is not hardwired to a specific choice of the (space) dimension d, and thus supports one-, two- and three-dimensional simulations. Note that DEAL.II-meshes are comprised of quadrilaterals and hexahedra instead of simplices, and can therefore technically not be referred to as "triangulations".

Although the algorithm itself looks relatively simple, even with the help of such a library its implementation already amounts to several thousand lines of code: Because DEAL.II only has a limited concept of "time", we need to provide our own functions and data structures that deal with space-time meshes and discretized functions (tuples like $(U^0, \dots, U^N)$) on such meshes. Finally, the solver itself must be written in such a way that it can be used to construct operators that relate to the direct problem.

In particular, it must be able to *efficiently* work with parameter tuples (c, ν, ρ, q) that arbitrarily combine discretized and continuous functions.[1]

Before we use this PDE-solver to solve inverse problems we should therefore prove that our implementation is correct. As is common, we accomplish this by solving a problem with a known classical solution $\bar{u}$, and then compare the error between the numerical- and exact solutions. This error should be small and, more importantly, should decrease when we refine the mesh. The theoretical framework for FEM even provides expectations for convergence rates. The time discretization using $\theta = 1/2$ (Crank-Nicolson) should be of second order in the maximal time step size

$$\Delta t := \max_{i=0,\dots,N-1} (t_{i+1} - t_i)$$

as long as the spatial discretization is fine enough and fixed, cf. [BGR10; GR05]. On the other hand, if we fix the discretization points in time and refine only in space, then we should see that the $L^2(I; L^2)$-error between $\bar{u}$ and its numerical approximation u should be of order h^2 in the maximal cell diameter $h > 0$. The same should hold for $\bar{u}'$ and ν, while $\|\bar{u} - u\|_{L^2(I;H_0^1)}$ should at least convergence linearly in h. In case of piecewise quadratic elements, these expectations increase to $\mathcal{O}(h^3)$ and $\mathcal{O}(h^2)$. These results already take into account that we neither have to deal with approximations of the domain's boundary $\partial\Omega$, nor the approximation of inhomogeneous boundary value functions.

Let $I := [0, 3]$, $\Omega := (0, \pi)^d$ and $d \in \{2, 3\}$. Of course we test our implementation with time- and space-dependent parameters; we choose

$$\begin{aligned} c(t,x) &:= \sqrt{1+t}(1+x_1) & \qquad q(t,x) &:= 1/20(t + x_1 x_2) \\ \rho(t,x) &:= (1+t)(1+x_2) & \qquad \nu(t,x) &:= 1/20\, |\cos(t x_2)| \end{aligned}$$

for $(t, x) \in I \times \Omega$ and would like the solution of the acoustic wave equation (5.14) to be equal to

$$\bar{u}(t,x) := \cos(t) \prod_{i=1}^{d} \sin(x_i).$$

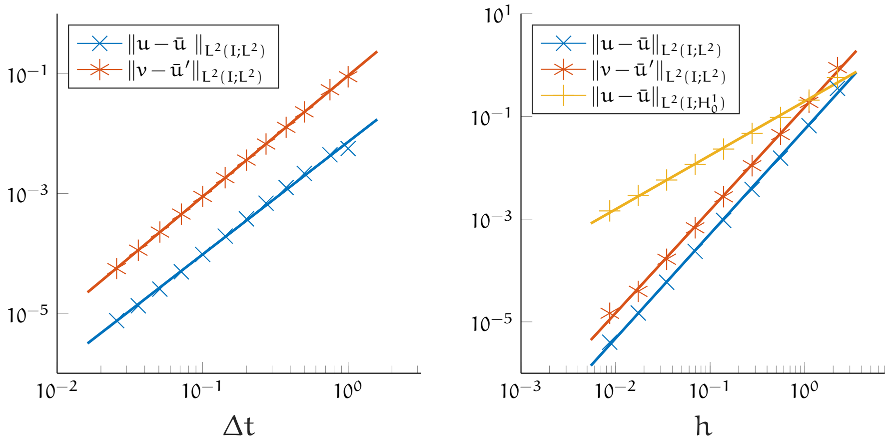
In order to achieve this, we define the initial values to be $u_0 := \bar{u}(0)$ and $\nu_0 := \bar{u}'(0) = 0$, and use the right-hand side

$$\begin{aligned} f(t,x) := {} & \frac{1}{\rho(t,x)} \frac{d}{dt}\left(\frac{\bar{u}'(t,x)}{c(t,x)^2}\right) - \operatorname{div}\left(\frac{\nabla \bar{u}(t,x)}{\rho(t,x)}\right) \\ & + q(t,x)\bar{u}(t,x) + \nu(t,x)\bar{u}'(t,x), \end{aligned}$$

which can be simplified by hand.

We employ the same uniform mesh of Ω in every time step, and distribute the time steps equidistantly as well. The linear systems in Algorithm 7.3 are solved iteratively, up to a relative tolerance of 10^{-8}

1 Typically, the parameter that is to be reconstructed will be discrete and the others will be continuous.



(a) successive refinement in time (b) successive refinement in space

Figure 7.2: Convergence of errors in the 2D reference scenario with piecewise linear elements

in the residual. For the entries of the needed matrices and right-hand side vectors we apply Gauß quadrature of fifth order. We also make use of this quadrature to compute the errors of the approximated u and v to the exact solution. A faster alternative would consist of looking at the differences to the FE-interpolations of $\bar{u}$ and $\bar{u}'$, but on very coarse meshes this could significantly distort the results.

We start by examining what happens if we refine only in time. For this, we consider piecewise linear elements, $d = 2$ and a fixed spatial mesh with $M = 263169$ degrees of freedom, which implies $h = \pi\sqrt{d}/(M^{1/d} - 1) \approx 8.7 \cdot 10^{-3}$. The resulting errors for various values of N are shown in Figure 7.2a. As expected, in this logarithmic presentation the $L^2(I; L^2(\Omega))$-errors of u and v describe straight lines with slopes 1.9 and 2.0 respectively. These observations match the theoretical convergence rate $\mathcal{O}(\Delta t^2)$.

Let us now fix the time discretization to $N = 256$ subintervals and vary the refinement in space. The corresponding results are depicted as Figure 7.2b and let us conclude

$$\|u - \bar{u}\|_{L^2(I;L^2(\Omega))} \approx \mathcal{O}(h^{2.0}), \quad \|v - \bar{u}'\|_{L^2(I;L^2(\Omega))} \approx \mathcal{O}(h^{2.0}),$$
$$\text{and} \quad \|u - \bar{u}\|_{L^2(I;H^1_0(\Omega))} \approx \mathcal{O}(h^{1.1}).$$

Again, these rates are on a par with the theoretical predictions. Experiments with quadratic elements and a higher $N = 1024$ yield the slopes 3.0, 2.9 and 2.0.

We have conducted analogous experiments in a three-dimensional setting and obtained the same results for the convergence in time and in space. With these observations we can safely assume that our implementation is correct enough to be used as the forward solver in the context of an inverse problem. However, before we close this section we would like to remark that Figure 7.2 has only limited use to draw conclusions about the approximation quality of our forward operators, because we had to choose a smooth solution $\bar{u}$ and smooth parameters. Reconstructed parameters are unlikely to be as regular, and thus might produce higher approximation errors in u.

7.2 INNER PRODUCTS

The use of the appropriate norms from the infinite-dimensional problem is (in the author's view) one of the main differences between the fields of inverse problems and optimization. Although it is much easier to just use p-norms of the numerical coefficient vectors, they probably do not fit to the continuous problem. Additionally, they are heavily dependent on the space-time mesh used for the discretization. This means that the ability to invert the problems (and the quality of this inversion) will also heavily depend on the mesh. In particular, we could not expect that the number of needed iteration steps is independent of the mesh, or even that it remains bounded for finer meshes. Moreover, all hyper-parameters have to be fine-tuned not with respect to the underlying problem, but with respect to the discretization.

Furthermore, the usual norms on $\mathbb{R}^n$ do not capture the smoothness in time of the parameter that is to be reconstructed. This smoothness is required by the theoretical results, and its absence might cause the inversion algorithm to diverge. Also, this coupling between the time steps of the reconstruction reduces the effective number of unknowns.

Naturally, there will always be some dependence on the mesh, for example due to approximation errors in the PDE solver, but the goal should be that the discretized problem approaches the continuous problem when the mesh is refined. Therefore, we will try to find numerical approximations of norms that are relevant to our problem. It is important to note that these norms do not only appear in an explicit way (when evaluating them) but also implicitly in the adjoint of the Fréchet-derivative of our forward operator F. Thus, it would be advantageous if we could gain access to transformation matrices that translate between transposed matrices and adjoints. To achieve this, we have to limit our focus to Hilbert spaces. In particular, we derive approximations for the $L^2(I;L^2)$-, $H^1(I;L^2)$- and $H^2(I;L^2)$ inner products and the corresponding transformation matrices, starting with $L^2(I;L^2)$.

7.2.1 *Approximation of* $L^2(I;L^2(\Omega))$

With $V_h \subset H_0^1(\Omega)$ we again denote our finite element space with basis $\{\varphi_1,\dots,\varphi_M\}$, and $0 = t_0 < \cdots < t_N = T$ describes the discretization in time. Suppose that we are given two functions $\tilde{u}$, $\tilde{v} \in C(I;V_h)$ and the task of evaluating $(\tilde{u},\tilde{v})_{L^2(I;L^2)}$. Numerically, we have access to $u = (u^0,\dots,u^N)$, $v = (v^0,\dots,v^N) \in \mathbb{R}^{M(N+1)}$, and the u^k, v^k are the coefficients of $u(t_k)$, $v(t_k)$ with respect to our chosen basis of V_h.

Of course this means that their L^2-inner product $(u(t_k), v(t_k))$ can be calculated without error via

$$(u(t_k), v(t_k)) = \sum_{i=1}^{M} \sum_{j=1}^{M} u_i^k v_j^k (\varphi_i, \varphi_j) = (u^k)^\mathsf{T} \mathbb{M} v^k,$$

where $\mathbb{M} \in \mathbb{R}^{M \times M}$ again denotes the mass matrix with entries (φ_i, φ_j). We propose that we approximate $L^2(I; L^2)$ in time using the trapezoidal rule, that is

$$\begin{aligned}(\tilde{u}, \tilde{v})_{L^2(I;L^2)} &\approx \sum_{k=0}^{N-1} \frac{t_{k+1} - t_k}{2} \left[(u(t_{k+1}), v(t_{k+1})) + (u(t_k), v(t_k)) \right] \\ &=: u^\mathsf{T}\, \mathbb{X}\, v =: (u, v)_{L^2(I;L^2)}. \end{aligned} \tag{7.12}$$

Had we chosen a space-time Galerkin ansatz with piecewise linear functions in time then this would even be exact. Equation (7.12) gives rise to a positive definite and symmetric matrix $\mathbb{X} \in \mathbb{R}^{M(N+1) \times M(N+1)}$, which means that the right-hand side of (7.12) defines an inner product on the coefficient vectors u and v. We abuse notation by denoting this inner product with $(u, v)_{L^2(I;L^2)}$.

The matrix $\mathbb{X}$ is not particularly useful for the evaluation of the inner product, and is probably too big to store in computer memory anyway. However, it becomes crucial if we require the adjoint (with respect to $L^2(I; L^2)$) of another linear operator, given as a matrix, based on the corresponding transposed matrix. Indeed, in this case $\mathbb{X}$ can be used as a transformation matrix since $(u, \mathbb{A}v)_{L^2(I;L^2)} = (\mathbb{A}^* u, v)_{L^2(I;L^2)}$ for all $\mathbb{A} \in \mathbb{R}^{M(N+1) \times M(N+1)}$ by means of $\mathbb{A}^* := \mathbb{X}^{-1} \mathbb{A}^\mathsf{T} \mathbb{X}$. Fortunately, $\mathbb{X}$ possesses a very simple structure; we have

$$\mathbb{X} = \begin{pmatrix} \frac{t_1 - t_0}{2}\mathbb{M} & & & & \\ & \frac{t_2 - t_0}{2}\mathbb{M} & & & \\ & & \ddots & & \\ & & & \frac{t_N - t_{N-2}}{2}\mathbb{M} & \\ & & & & \frac{t_N - t_{N-1}}{2}\mathbb{M} \end{pmatrix}.$$

Solving linear systems with this block-diagonal matrix simply amounts to solving $N+1$ systems of size M. Furthermore, this matrix can be easily adapted to the case of time-dependent meshes. It is also easily generalized to other norms in the spatial variable. If we want to use $H_0^1(\Omega)$ instead, then all occurrences of the mass matrix $\mathbb{M}$ have to be replaced by a matrix with entries $(\varphi_i, \varphi_j)_{H_0^1} = (\varphi_i, \varphi_j) + (\nabla\varphi_i, \nabla\varphi_j)$. Note that this is the system matrix of the finite element discretization of the elliptic equation $\varphi - \Delta\varphi = g$ with homogeneous Dirichlet boundary conditions. Thus, in this case the application of $\mathbb{X}^{-1}$ has a smoothening effect on the spatial variable.

7.2.2 *Including time derivatives*

The inner product of $L^2(I; L^2(\Omega))$ is not enough to ensure differentiability of the reconstructed parameters. This might not pose a problem for q and ν, but it will probably cause the reconstruction of the mass density ρ or the wave speed c to diverge. From the theoretical results we know that the forward operator F is only differentiable in those ρ and c, that are at least three times weakly differentiable and possess essentially bounded derivatives. We will only approximate $H^k(I; L^2(\Omega))$ for $k = 1, 2$, with the hope that this will be sufficient. An even higher k would lead to excessive computational complexity in the linear systems that have to be solved. The large computing times are also the main reason why we will resort to simple difference quotients and not higher order approximations of derivatives (e.g. based on spline-interpolation). The other reason is the fact that our forward solver already has an approximation error of order $\mathcal{O}(\Delta t^2)$. An assortment of difference quotients can be found in [For88]. We approximate $\tilde{u}'(t_i)$ with simple central difference quotients inside I and one-sided quotients at the boundary, i.e.

$$\tilde{u}'(t_i) \approx \frac{\tilde{u}(t_{i+1}) - \tilde{u}(t_{i-1})}{t_{i+1} - t_{i-1}}, \quad i = 1, \ldots, N-1,$$
$$\tilde{u}'(t_0) \approx \frac{\tilde{u}(t_1) - \tilde{u}(t_0)}{t_1 - t_0} \;\text{ and }\; \tilde{u}'(t_N) \approx \frac{\tilde{u}(t_N) - \tilde{u}(t_{N-1})}{t_N - t_{N-1}}.$$

This approximation is of second order inside the time domain and of first order at the two boundary nodes. Naturally, this setup implies that the approximation of $\tilde{u}'(t_i)$ belongs to V_h as well, and its coefficients, which we will denote by $u' \in \mathbb{R}^{M(N+1)}$ (the prime is only a notation in this case) can easily be calculated from u. In fact, it is given by $\mathbb{D}_{(1)}u$ with a tridiagonal block matrix $\mathbb{D}_{(1)}$. Furthermore, each block of this matrix is a multiple of the $M \times M$ unit matrix.

For the second derivatives we employ the common second-order formula

$$\tilde{u}''(t_i) \approx 4\,\frac{u(t_{i-1}) - 2u(t_i) + u(t_{i+1})}{(t_{i+1} - t_{i-1})^2}.$$

At the boundary nodes we simply copy the values at their neighbors, $\tilde{u}''(t_0) \approx \tilde{u}''(t_1)$ and $\tilde{u}''(t_{N+1}) \approx \tilde{u}''(t_{N-1})$. For three times differentiable u this produces a first-order approximation. Again, this gives rise to $u'' = \mathbb{D}_{(2)}u$ with an additional matrix $\mathbb{D}_{(2)}$ that has a similar structure as $\mathbb{D}_{(1)}$. Of course we approximate $\tilde{v}'$ and $\tilde{v}''$ in the same way.

We would like to have the option of giving different importance to the parts of the H^2 inner product. Thus, we introduce weights $\alpha, \beta \geqslant 0$ and set

$$(\tilde{u}, \tilde{v})_{H^2(I;L^2)} := (\tilde{u}, \tilde{v})_{L^2(I;L^2)} + \alpha \left(\tilde{u}', \tilde{v}'\right)_{L^2(I;L^2)} + \beta \left(\tilde{u}'', \tilde{v}''\right)_{L^2(I;L^2)}.$$

As long as $\beta > 0$ (even if $\alpha = 0$) this is still equivalent to the standard H^2-inner product (highest and lowest order derivatives are sufficient, cf. [AF03, Theorem 5.2]), but will produce different numerical reconstructions when used in the context of inverse problems. Obviously we do not need to look at the H^1-case separately; it is included as the special case $\beta = 0$, $\alpha > 0$. With the notation of the previous subsection we have $(\tilde{u}', \tilde{v}')_{L^2(I;L^2)} \approx \left(\mathbb{D}_{(1)}u, \mathbb{D}_{(1)}v\right)_{L^2(I;L^2)}$ and $(\tilde{u}'', \tilde{v}'')_{L^2(I;L^2)} \approx \left(\mathbb{D}_{(2)}u, \mathbb{D}_{(2)}v\right)_{L^2(I;L^2)}$, thus

$$
\begin{aligned}
(\tilde{u}, \tilde{v})_{H^2(I;L^2)} &\approx u^T \mathbb{X} v + \alpha \left(\mathbb{D}_{(1)}u\right)^T \mathbb{X}\mathbb{D}_{(1)}v + \beta \left(\mathbb{D}_{(2)}u\right)^T \mathbb{X}\mathbb{D}_{(2)}v \\
&= u^T \left(\mathbb{X} + \alpha \mathbb{D}_{(1)}^T \mathbb{X}\mathbb{D}_{(1)} + \beta \mathbb{D}_{(2)}^T \mathbb{X}\mathbb{D}_{(2)}\right) v.
\end{aligned}
$$

The key fact, that allows the efficient inversion of this huge matrix is that $\mathbb{X}$ commutes with the derivative matrices. Hence,

$$
(\tilde{u}, \tilde{v})_{H^2(I;L^2)} \approx u^T \left(\mathbb{I} + \alpha \mathbb{D}_{(1)}^T \mathbb{D}_{(1)} + \beta \mathbb{D}_{(2)}^T \mathbb{D}_{(2)}\right) \mathbb{X} v.
$$

We have already discussed that $\mathbb{X}^{-1}$ can be evaluated by solving $N+1$ linear systems of size M. The other matrix consists only of multiples of $M \times M$ unit matrices, and can therefore be inverted by solving M linear systems of size $N+1$. For $d > 1$ it is safe to assume $N + 1 \ll M$, thus even a dense matrix of size $M \times M$ will not cause memory problems. However, since it has to be used to solve *a lot* of linear systems, it is advised to precompute a factorization (e.g. LU or Cholesky) of it. In our implementation we utilize the UMFPACK library [Dav04] to perform a sparse LU factorization.

We would like to make two remarks. First, this approach requires time-independent meshes. Without this assumption, we would not be able to easily decompose the linear system of size $\sum_{i=0}^{N} M_i$ into a number of systems of size $N + 1$. Second, we would like to note that the time components of these matrices are very similar to the ODEs we would have to solve to obtain the adjoint of the embedding $H^2(I) \hookrightarrow L^2(I)$.

7.3 DISCRETIZED OPERATORS

Using the PDE solver of section 7.1, implementing approximations of the forward operator $F\colon (c, \nu, \rho, q) \mapsto u$ is not difficult. To avoid the runtime cost of interpolation operators, we discretize the parameters on the same mesh and with the same finite element basis (piecewise linear in space) as the solution u.

The same holds for ∂F, which (according to Theorem 5.4) involves the solution of the same PDE with more complicated right-hand sides. These terms can be computed using quadrature; for the additional time derivative in the right-hand side of $\partial_c F$ we additionally employ a simple finite-difference scheme. Should this cause oscillations in the reconstruction, then one should consider more stable approaches instead, for example spline interpolation [HS01] or higher order finite differences that introduce more coupling between nodes.

Most inversion algorithms require access to the adjoint of ∂F in order to map back to the parameter space. This is the reason why we already

$$\begin{tikzcd}
 & {[\partial F(x)]_d} \arrow[r] & {\left([\partial F(x)]_d\right)^*} \\
\partial F(x) \arrow[ur, bend left] \arrow[dr, bend right] & & \neq \\
 & \partial F(x)^* \arrow[r] & {[\partial F(x)^*]_d}
\end{tikzcd}$$

Figure 7.3: Should we use the approximation of the adjoint or the adjoint of the approximation?

spent some effort on the adjoints of ∂S and ∂F in sections 4.2.3 and 5.3.1, which resulted in the adjoint problem (5.24). We also know that if ν and ρ' are time-independent then the adjoint equation has the same structure as the original PDE. It only has to be solved backwards in time. For the highly interesting case of a nonlinear (with respect to time) mass density we require a specialized solver for the adjoint problem. We could go back to section 7.1 and derive such a solver. However, even in this case we will only obtain an approximation $[(\partial F(x)^*)]_d$ of $\partial F(x)^*$, which will not be equal to the adjoint of the discretized $\partial F(x)$, even if the correct discretized norms are used (as visualized in Figure 7.3). Although this approximation error will decrease with $\mathcal{O}(h^2 + \Delta t^2)$, it can cause serious complications in the solution of linear inverse problems with $\partial F(x)$ as a forward operator. For the solution of such problems with the linear Landweber method this effect has already been analyzed on a theoretical basis, see [EH18]. In numerical experiments we have observed that this can also happen with the conjugate gradient method, especially near the end of the reconstruction process, when the linear problems become harder to solve. In the sequel, we will devise an algorithm that should produce the *matrix transpose* (up to the tolerances of the linear solvers) of the discretized $\partial F(x)$ instead. The matrices from the previous section can then be used to transform this transposed matrix to the desired inner product.

Before we start, we would like to emphasize that one could also argue *in favor* of the analytical adjoint. It is more likely to capture the nature (e.g. smoothness) of functions in the domain of definition of ∂F. In contrast, the transpose of the discretized derivative might put more focus on effects that relate to the discretization. If its usage does not cause problems, one should therefore prefer the analytical adjoint.

As outlined above, we discretized the partial derivatives of F by inserting suitable right-hand side vectors into Algorithm 7.3, which results in the vector $U = (U^0, \dots, U^N)$. The linearized wave fields possess homogeneous initial values, hence the algorithm is linear in the right-hand side vector $\bar{F} := (F^0, \dots, F^N)$. In consequence, it has to possess a representation as a $M(N+1) \times M(N+1)$-matrix L with $L\bar{F} = U$. Due to its size, and the fact that it involves matrix inversions, we cannot simply assemble L and then transpose it. Instead, we will derive a way to evaluate it that is similar to Algorithm 7.3.

From the system (7.11) that has to be solved in each time step, we conclude that L can be decomposed into

$$L = EX^{-1}R, \tag{7.13}$$

where the matrix R is of size $2(N+1)M \times (N+1)M$ and performs a suitable weighting of $\bar{F}$ such that the right-hand sides of (7.11) emerge. The matrix X contains the matrices that have to be solved in each step to acquire U^{n+1} and V^{n+1}, starting with U^0 and V^0. Thus, it is a square matrix of size $2(N+1)M \times 2(N+1)M$. Finally, since we only want to keep U, the $(N+1)M \times 2(N+1)M$ matrix E simply throws away every second M-block of its argument. This implies that it has the shape

$$E = \begin{pmatrix} 1 & 0 & & & & & & \\ & & 1 & 0 & & & & \\ & & & & 1 & 0 & & \\ & & & & & \ddots & \ddots & \\ & & & & & & 1 & 0 \end{pmatrix},$$

where each number represents the corresponding multiple of the $M \times M$ identity matrix.

The entries of the matrices X and R are not so easily found. We begin by having a closer look at the matrix R. The initial conditions $U^0 = V^0 = 0$ do not make use of $\bar{F}$, hence the first $2M$ rows of R must be zero. Every subsequent block of $2M$ rows is determined by (7.11). The right-hand side of (7.11a) depends on $\theta^2 F^{n+1} + \theta(1-\theta)\mathbb{D}^{n,n+1}\mathbb{M}^{-1}F^n$, therefore the first M rows of such a block perform this addition of the entries $Mn+1, \dots, M(n+1)$ and $M(n-1)+1, \dots, Mn$ of the input. The second M rows are given by the equation (7.11b) and combine θF^{n+1} with $(1-\theta)\mathbb{D}^{n,n+1}\mathbb{M}^{-1}F^n$. We conclude that the matrix R reads

$$R = \begin{pmatrix} \boxed{\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}} & & \\ \boxed{\begin{matrix} \theta(1-\theta)\mathbb{D}^{0,1}\mathbb{M}^{-1} & \theta^2 \\ (1-\theta)\mathbb{D}^{0,1}\mathbb{M}^{-1} & \theta \end{matrix}} & & \\ & \ddots \quad \ddots & \\ & & \boxed{\begin{matrix} \theta(1-\theta)\mathbb{D}^{N-1,N}\mathbb{M}^{-1} & \theta^2 \\ (1-\theta)\mathbb{D}^{N-1,N}\mathbb{M}^{-1} & \theta \end{matrix}} \end{pmatrix}.$$

Again, each entry in this matrix is to be understood as an $M \times M$ block. Note that the application of R does not have to involve the solution of linear systems if $\bar{F}$ consists of inner products with elements of V_h, which cancel with the inverse of the mass matrix.

Now we turn to the matrix X. Its inverse has to map $y = (y^0, \dots, y^N) \coloneqq R\bar{F}$ onto the solution vector $w \coloneqq \left(U^0, V^0, \dots, U^N, V^N\right)^T$. By design of R, the first 2M entries y^0 of y are zero, which is what we want (U^0, V^0) to be as well. Thus, the first rows of X can be set to $\mathbb{I}_{2M\times 2M}$. For $n = 0, \dots, N$, the tuple $(U^{n+1}, V^{n+1}) \in \mathbb{R}^{2M}$ is the solution of a system that has a linear combination of y^{n+1}, U^n and V^n as the right-hand side. Therefore, there have to exist $2M \times 2M$-matrices $X^{(1)}_{n+1}$, $X^{(2)}_{n+1}$ such that

$$X^{(2)}_{n+1}\begin{pmatrix} U^{n+1} \\ V^{n+1} \end{pmatrix} = y^n - X^{(1)}_{n+1}\begin{pmatrix} U^n \\ V^n \end{pmatrix}. \tag{7.14}$$

The entries of these matrices are of course again given by (7.11) and are equal to

$$X^{(1)}_{n+1} = \begin{pmatrix} \theta(1-\theta)\mathbb{D}^{n,n+1}\mathbb{M}^{-1}\mathbb{A}^n - k_{n+1}(k_{n+1}\mathbb{C}^{n+1} + \theta\mathbb{B}^{n+1}) & \theta(1-\theta)(\mathbb{D}^{n,n+1}\mathbb{M}^{-1}\mathbb{B}^n - \mathbb{B}^{n+1}) - k_{n+1}(\theta\mathbb{C}^{n,n+1} + (1-\theta)\mathbb{C}^{n+1}) \\ (1-\theta)\mathbb{D}^{n,n+1}\mathbb{M}^{-1}\mathbb{A}^n & -k_{n+1}\mathbb{C}^{n,n+1} + (1-\theta)\mathbb{D}^{n,n+1}\mathbb{M}^{-1}\mathbb{B}^n \end{pmatrix}$$

$$X^{(2)}_{n+1} = \begin{pmatrix} k^2_{n+1}\mathbb{C}^{n+1} + k_{n+1}\theta\mathbb{B}^{n+1} + \theta^2\mathbb{A}^{n+1} & 0 \\ \theta\mathbb{A}^{n+1} & k_{n+1}\mathbb{C}^{n+1} + \theta\mathbb{B}^{n+1} \end{pmatrix}.$$

Finally, we can place the systems (7.14) in the matrix X to conclude that is takes the form

$$X = \begin{pmatrix} \boxed{\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}} & & & & \\ X^{(1)}_1 & X^{(2)}_1 & & & \\ & X^{(1)}_2 & X^{(2)}_2 & & \\ & & \ddots & \ddots & \\ & & & X^{(1)}_N & X^{(2)}_N \end{pmatrix}.$$

From the decomposition (7.13) of L we immediately obtain

$$L^T = R^T(X^T)^{-1}E^T.$$

The application of E^T involves intercalating M zeros between each M-block of the argument to become a vector of length $2M(N+1)$, i.e. $E^T g = (g_0, 0, g_1, 0, \dots, g_N, 0)^T$ for $g \in \mathbb{R}^{M(N+1)}$. The matrix X is a lower triangular block matrix, thus X^T is an upper triangular block matrix. By backward substitution, its inversion can be reduced to solving $N+1$ systems of size 2M. This is reminiscent of the theoretical adjoint that also involved solving the wave equation backwards in time. The subsequent application of R^T is easily accomplished, and can even be incorporated into the inversion of X^T in order to avoid having to store (or worse, reassemble) the matrices $\mathbb{D}^{n,n+1}$.

Due to the structure of the right-hand side $E^T g$ and the blocks of X^T we can reduce the computational effort for each time step $i = N, \dots, 0$ to the solution of a maximum of two linear systems of size $M \times M$:

- $i = N$: The last $2M$ equations of $X^T w = E^T g$ read

$$X_N^{(2)\,T} \begin{pmatrix} U^N \\ V^N \end{pmatrix} = \begin{pmatrix} g^N \\ 0 \end{pmatrix}.$$

  Since $\left(X_N^{(2)}\right)_{12} = 0$, this breaks down into $V^N = 0$ and the solution of $\left(X_N^{(2)}\right)_{11} U^N = g^N$.

- $i = N-1, \dots, 1$: Here we have to solve

$$X_i^{(2)\,T} \begin{pmatrix} U^i \\ V^i \end{pmatrix} + X_{i+1}^{(1)\,T} \begin{pmatrix} U^{i+1} \\ V^{i+1} \end{pmatrix} = \begin{pmatrix} g^i \\ 0 \end{pmatrix}.$$

  Keeping in mind that $X_i^{(2)\,T}$ is a lower triangular block matrix, this system can first be solved for V^i, and then for U^i.

- $i = 0$: The first $2M$ equations read

$$\begin{pmatrix} U^0 \\ V^0 \end{pmatrix} + X_1^{(1)\,T} \begin{pmatrix} U^1 \\ V^1 \end{pmatrix} = \begin{pmatrix} g^0 \\ 0 \end{pmatrix}.$$

  They do not require the solution of any linear systems.

Even in the case of time-dependent meshes we could therefore write our scheme for the evaluation of L^T as an algorithm similar to Algorithm 7.3, but with the reversed order of time steps. To save memory, it is again advisable to first calculate all quantities that depend on time step $i+1$, clear the "old" matrices and only then assemble the matrices that involve time step i.

As we have noted in the introduction of this section, the solution of the wave equation is only one part in the evaluation of ∂F, because the right-hand side depends on the linearization direction. In consequence, the adjoint has to apply the corresponding "transposed" operations *after* L^T. However, they are relatively easy to derive, since the original operators mainly consist of multiplications of nodal values and finite differences in the time domain. Thus, we forgo their presentation here.

NUMERICAL COMPARISONS Now we possess two solvers that approximate the adjoint of the operator $\widehat{L} \in \mathcal{L}\left(L^2(I; L^2(\Omega))\right)$, which maps functions f to the solutions u of the acoustic wave equation with f as the right-hand side. On the one hand, we can use the PDE solver from

N	$L^* =$ Backwards-Solving		$L^* = L^T$	
	trivial	complex	trivial	complex
8	$7.69 \cdot 10^{-2}$	$1.63 \cdot 10^{-2}$	$6.63 \cdot 10^{-9}$	$1.83 \cdot 10^{-9}$
16	$9.82 \cdot 10^{-2}$	$3.04 \cdot 10^{-1}$	$1.07 \cdot 10^{-8}$	$7.42 \cdot 10^{-9}$
32	$3.17 \cdot 10^{-2}$	$8.63 \cdot 10^{-3}$	$6.95 \cdot 10^{-10}$	$1.44 \cdot 10^{-9}$
64	$8.19 \cdot 10^{-3}$	$2.11 \cdot 10^{-2}$	$1.41 \cdot 10^{-9}$	$3.04 \cdot 10^{-10}$
128	$5.76 \cdot 10^{-3}$	$1.52 \cdot 10^{-1}$	$2.63 \cdot 10^{-9}$	$2.43 \cdot 10^{-9}$
256	$2.64 \cdot 10^{-3}$	$3.25 \cdot 10^{-3}$	$8.85 \cdot 10^{-10}$	$9.56 \cdot 10^{-10}$
512	$1.09 \cdot 10^{-3}$	$1.05 \cdot 10^{-1}$	$5.47 \cdot 10^{-9}$	$8.78 \cdot 10^{-10}$

Table 7.1: Values of Δ_{adj} in 2D with two different sets of parameters

Algorithm 7.3 and run it backwards in time, keeping in mind that this only yields the correct operator if $B_{c,\nu,\rho} = 0$, i.e. if ν vanishes and ρ depends linearly on time. On the other hand, we are able to transform the transpose L^T of the discretized $\widehat{L}$ to the correct inner products in definition- and image space by making use of the preceding section's transformation matrices.

We would like to give a short numerical comparison between these two methods by analyzing the average relative error

$$\Delta_{\mathrm{adj}} := \frac{1}{K} \sum_{i=1}^{K} \frac{\left|(f_i, L^* g_i)_{L^2(I;L^2)} - (Lf_i, g_i)_{L^2(I;L^2)}\right|}{\left|(Lf_i, g_i)_{L^2(I;L^2)}\right|}$$

over a sample size of $K = 10$. The vectors $f_i, g_i \in \mathbb{R}^{M(N+1)}$ $(i = 1, \dots, K)$ are initialized with white noise and then slightly smoothed in time and space.

For this experiment we consider two sets of parameters: First, the trivial acoustic wave equation $u'' - \Delta u = f$, that is $c = \rho = 1$ and $q = \nu = 0$. Second, the more complex parameters

$$\begin{aligned} c(t,x) &:= x_1 \cos(t) + 3/2, & \nu(t,x) &:= |x_2| \sin(t), \\ \rho(t,x) &:= \|x\| + \sin(t) + 3/2, & q(t,x) &:= \mathbb{1}_{B(0,1/2)}(x) \sin(\pi t). \end{aligned}$$

In both cases, the space-time domain is given by $\Omega := [-1,1]^d$ and the time interval is $I := [0,3]$.

We ran this test with $d = 2$ for different time discretizations, while keeping the spatial mesh fixed at $M = 1089$ degrees of freedom. This resulted in the values for Δ_{adj} as seen in Table 7.1. By setting $d = 3$ we obtained very similar results.

As predicted, for complex parameters backward-integration does not perform well at all, thus we should not rely on this method for the inverse problem. For trivial parameters the errors start high (in the percent range), before they decrease to a few per mill for very fine discretizations. If we instead consider the solver based on matrix transposes, the errors become almost negligible for both sets of parameters. This is as expected, since we do not have to deal with any approximation errors here, even the calculation of the inner products exactly matches the transformation matrices. However, since the implementations of L and L^T contain iterative solvers, the errors are still magnitudes larger than machine accuracy ($\approx 10^{-16}$). Instead, they approximately possess the same magnitude as the relative tolerance for the linear solvers, which we had set to 10^{-8}.

8 NUMERICAL INVERSION OF THE ACOUSTIC WAVE EQUATION

We are still dealing with the acoustic wave equation in form of (5.15), which we repeat once again for the reader's convenience; it is

$$\begin{aligned} \left(\frac{1}{\rho c^2}u'\right)' + \left(\frac{\rho'}{\rho^2 c^2} + \nu\right)u' - \operatorname{div}\frac{\nabla u}{\rho} + qu = f \qquad & \text{in } (0,T)\times\Omega, \\ u(0) = \left(c^{-2}\rho^{-1}u'\right)(0) = 0 \qquad & \text{in } \Omega, \\ u = 0 \qquad & \text{on } (0,T)\times\partial\Omega. \end{aligned}$$

In Chapter 5 we have built the operator $F\colon D(F)\cap W^{(k)} \to Y^{(k)}$, which maps the parameters c, ν, ρ and q to the weak solution u of this problem, and we have also analyzed F on a theoretical level. The framework of Chapter 7 yields data structures for the discretized parameters and approximations of F and its Fréchet-derivative. Thus, we should finally be able to tackle the actual reconstruction of parameters.

Note that we will only consider the identification of exactly one of the parameters, while the others are known and fixed. This allows us to discover which of the four parameters causes the most difficulties in the reconstruction. Furthermore, it prevents the inverse problems from becoming hopelessly under-determined.

We will start off this chapter by a discussion of possible regularization methods, and in which function spaces we are going to apply them. In Section 8.2 we present two example parameters that we will be using as "ground truths" throughout the sequel. Indeed, their first use directly follows; in Section 8.3 we analyze the behavior of the error in the reconstruction of these parameters when the noise level in the data decreases. An important aspect for possible applications will be covered in Section 8.4, namely the reconstruction from "partial" knowledge of u. This becomes important for example if one is only able to observe the wave field near the boundary of Ω. There, we also make some experiments with time-independent parameters, and how knowing whether the ground-truth is static can affect the reconstruction quality.

8.1 REGULARIZATION APPROACH

We have formulated the operator F to act on the open subset $D(F)$ of the Banach space $W^{(k)} \subset L^\infty(I;L^\infty(\Omega))^4$ (as defined in equations (5.19) and (5.20) on page 78) and produce values in $Y^{(k)} \subset L^\infty(I;L^2(\Omega))$. To obtain a well-defined F we may use $k=0$, but for differentiability in the parameters c and ρ we have to use at least $k=2$. We directly replace the image space $Y^{(k)}$ by $L^2(I;L^2(\Omega))$. Otherwise, we would implicitly assume differentiability of (measurement-) noise, like we already discussed in Section 4.2.3.

As pointed out in the introduction, we will only consider the reconstruction of one of the four functions c, ν, ρ and q. This means that we need to split F into four parts. Furthermore, for the numerical experiments it is more convenient to write the searched for parameter as the sum of a known, smooth "background" function $c_b, \nu_b, \rho_b, q_b \colon I \times \Omega \to \mathbb{R}$ and a perturbation, and only reconstruct this perturbation. For the inversion, this has the same effect as using the background parameter as the initial guess (or shifting penalty terms by this function), and we may recover the original operator by setting this background to zero. Nevertheless, this small modification makes the reconstruction error easier to interpret: If for example $\|c_b\|$ is large compared to the sought for perturbation and we initialize with c_b (we must start with a positive function anyway), then the computed reconstruction error will immediately be small, because it is tainted by this a-priori information.

We note that we can write $D(F)$ as the cartesian product

$$D(F) = D(F_c) \times L^\infty(I; L^\infty(\Omega)) \times D(F_\rho) \times L^\infty(I; L^\infty(\Omega))$$

to separate the constraints on c and ρ into the two sets

$$D(F_c) := \Big\{ c \in W^{1,\infty}(I; L^\infty(\Omega)) \;\Big|\; c_0 + \delta \leqslant c_b + c \leqslant c_1 - \delta \text{ a.e. in } I \times \Omega \text{ for a } \delta > 0 \Big\}, \tag{8.1}$$

$$D(F_\rho) := \Big\{ \rho \in W^{1,\infty}(I; L^\infty(\Omega)) \;\Big|\; \rho_0 + \delta \leqslant \rho_b + \rho \leqslant \rho_1 - \delta \text{ a.e. in } I \times \Omega \text{ for a } \delta > 0 \Big\}. \tag{8.2}$$

By fixing exactly three of the four parameters, and choosing k as small as possible, we obtain differentiable forward operators

$$\begin{aligned}
&F_c \colon D(F_c) \cap W^{3,\infty}(I; L^\infty(\Omega)) \to L^2(I; L^2(\Omega)), && F_c := F(c_b + \cdot, \nu_b, \rho_b, q_b)\\
&F_\nu \colon L^\infty(I; L^\infty(\Omega)) \to L^2(I; L^2(\Omega)), && F_\nu := F(c_b, \nu_b + \cdot, \rho_b, q_b),\\
&F_\rho \colon D(F_\rho) \cap W^{3,\infty}(I; L^\infty(\Omega)) \to L^2(I; L^2(\Omega)), && F_\rho := F(c_b, \nu_b, \rho_b + \cdot, q_b),\\
&F_q \colon L^\infty(I; L^\infty(\Omega)) \to L^2(I; L^2(\Omega)), && F_q := F(c_b, \nu_b, \rho_b, q_b + \cdot).
\end{aligned}$$

We will always implicitly assume that the fixed parameters possess the appropriate smoothness. For example, if we analyze F_c then q_b, $\nu_b \in W^{2,\infty}(I; L^\infty(\Omega))$ and $\rho_b \in D(F_\rho) \cap W^{3,\infty}(I; L^\infty(\Omega))$ must hold.

Getting our numerical experiments to conform to this theory is difficult because in inverse problems we like to work with Hilbert spaces, which is easily seen on the standard literature (like [EHN00; Kir11; Rie03]). The discipline was founded on Hilbert space theory, and this only changed very recently, when researchers recognized the importance of Banach spaces for problems that stem from applications or involve sparsity constraints. Since then, methods like Tikhonov regularization and some iterative (Newton-) techniques have become available in Banach space settings. We refer to [SLS06; Sch+12] for an overview of some of these methods, and to [Mar18; Wal18] for some very recent developments. Unfortunately, all of these methods require some assumptions about the structure of the involved spaces, which then imply that the domain of definition must belong to a *reflexive* Banach space. Obviously this is not the case for our problem, since we have to deal with L^∞-type spaces in both the space- and time variables. We could simply replace these spaces by the corresponding L^p-space for a large p, in the hope that this approximates the topology of L^∞ well enough, but of course this would still not match our theoretical results. Furthermore, in many applications one is more interested in heavily localized, *sparse* parameters that are obtained by choosing p on the other end of the spectrum, i.e. close to 1. The L^∞-norm is agnostic to more "spread out", slowly-varying parameters. The best compromise seems to be to use $p = 2$ and stay in the Hilbert space framework, which also (typically) involves less implementation- and computational effort.

8.1.1 *Reformulating the problem in Hilbert spaces*

There are two ways of reshaping the definition spaces into Hilbert spaces. On the one hand, we can try to find a Hilbert space Z such that $Z \subset L^\infty(I; L^\infty)$ or $W^{3,\infty}(I; L^\infty)$, depending on the parameter. In the time variable this is quickly achieved since $H^{m+1}(I; L^\infty) \subset W^{m,\infty}(I; L^\infty)$ holds for $m \in \mathbb{N}_0$ (cf. Theorem A.10). For $L^\infty(\Omega)$ an analogous argument can be made, but it is dependent on the space dimension d: In 1D we see $H^l(\Omega) \subset L^\infty(\Omega)$ for $l \geqslant 1$ (special case of Theorem A.10 for $X = \mathbb{R}$). In two- or three-dimensional scenarios this holds as long as $l \geqslant 2$ (Sobolev imbedding theorem, cf. [AF03]). Consequently, we may replace $W^{m,\infty}(I; L^\infty)$ by $H^{m+1}(I; H^2(\Omega))$. This approach has the advantage of still conforming to the theory and pointwise constraints yield open subsets of these spaces. However, for c and ρ we have $m = 3$, and would have to deal with functions that are four times weakly differentiable in time with values in $H^2(\Omega)$. In particular, this implies that both the function and its weak derivatives up to fourth order are twice weakly differentiable in space. Enforcing this much regularity of the numerical solution requires a lot of effort; in fact, our finite element basis does not provide it. More importantly, from an applications viewpoint this requirement seems excessive, if not absurd.

The other way around this problem is to simply use the inner product we wish to use anyway, and pretend that the resulting operators are still well-defined and differentiable. We can try to justify this by restricting our definition spaces to some subspace E with an arbitrary huge, but finite dimension. There, all norms are equivalent, and we may try to use the $H^2(I; L^2)$-, $H^1(I; L^2)$- or even $L^2(I; L^2)$-norms instead of $W^{k,\infty}(I; L^\infty)$. If we take this subspace to be an incredibly fine finite element space, then our discretized operators will still approximate the corresponding problem for all practically achievable mesh refinements. For example, let

V_h^∞ denote the space of piecewise linear Lagrange elements on a mesh of Ω after it has been globally refined 100 times, which means that it contains at least 2^{100d} cells. We can proceed in a similar way for the time variable, e.g. by uniformly partitioning I into 2^{100} subintervals. Then we can define E to be the space of all splines $I \to V_h^\infty$ of order four, and imagine that we work with the restrictions

$$F_c\big|_E,\ F_\nu\big|_E,\ F_\rho\big|_E \text{ and } F_q\big|_E.$$

Of course this has no effect on the numerical computations, and only serves the purpose of making us feel better about the fact that we ignore crucial parts of the theoretical analysis.

On paper, the derivatives of these restricted operators cannot be ill-posed anymore, because they possess a finite-dimensional range (as is always the case for the discretized operators). However, they will probably turn out to be ill-conditioned.

8.1.2 *Regularization methods*

We assume that the reader is familiar with the concept of inverse problems and regularization. The goal of this subsection is to justify why we chose a particular regularization scheme, give a reminder on how it works and discuss our modifications to the relevant algorithms.

No matter which parameter we want to reconstruct, we have to deal with a nonlinear operator $G\colon D(G) \subset X \to Y$, which maps between some Hilbert spaces X and Y. We assume that both G and its derivatives are ill-posed (or at least ill-conditioned), and want to solve the equation

$$Gx = y$$

for an approximation of $x \in D(G)$. However, we only have access to a noisy version $y^\varepsilon \in Y$ of $y \in Y$. The relative noise level $\varepsilon = \|y^\varepsilon - y\|_Y / \|y\|_Y$ is assumed to be known. This means that we may use the discrepancy principle: In order to suppress unwanted effects due to the noise, we are satisfied as soon as we have found $x^\varepsilon \in D(G)$ such that

$$\|Gx^\varepsilon - y^\varepsilon\|_Y \leqslant \tau\varepsilon\|y\|_Y \approx \tau\varepsilon\|y^\varepsilon\|_Y \tag{8.3}$$

is fulfilled for a $\tau > 1$ that was chosen in advance.

Without doubt, the most popular regularization method is Tikhonov regularization. It is very flexible because a-priori information can easily be integrated into the penalty term. However, it only takes care of the ill-posedness, afterwards one still needs a solver for nonlinear optimization problems. This again limits the choice of norms and penalty terms, because this solver needs to be able to deal with their structure. For example, classical optimization approaches (eg. SQP-methods as employed in the solver WORHP [BW13]) require the functional to be differentiable. If the functional is at least convex, then the optimization problem can be solved by primal-dual algorithms like the one proposed by CHAMBOLLE and Pock [CP11], which has gained immense popularity within the imaging community. At least in the author's personal experience, Tikhonov regularization results in large computing times. Perhaps this is due to the fact that we have limited ways of re-using the solution of the problem for a different regularization parameter, other than using it as the initial guess for the next regularization parameter on the list.

Alternatives to Tikhonov are iterative methods like the nonlinear Landweber method, which updates the current estimate $x_n \in X$ (starting with some $x_0 \in D(G)$) with one step of the linear Landweber method applied to the linearized equation

$$G'(x_n)s = y^\varepsilon - G(x_n) =: b_n^\delta, \tag{8.4}$$

that is $x_{n+1} := x_n - \omega_n G'(x_n)^*(G(x_n) - y^\varepsilon)$. It is known to be slow and it has an additional scalar parameter $\omega_n > 0$ that has to be guessed (the theoretical bounds involve the operator norm of $G'(x_n)$ and are therefore typically useless in practice) correctly in order for the method to converge.

NEWTON-METHODS By updating x_n with the solution $s \in X$ of the linearized equation (8.4) we arrive at the class of Newton methods. However, we know that these linearized problems are ill-posed as well, which makes their exact solution impossible. Hence, we must be satisfied with an "inexact" approximation of s, obtained by any regularization method for linear inverse problems. Unfortunately, even if ε is known, the noise δ in the right-hand side b_n^δ of the linearized equation is not, because it also contains the linearization error. Thus, we cannot rely on the discrepancy principle to find a suitable regularization parameter, and must figure out a relative tolerance $\mu_n \in [0, 1]$ to substitute the discrepancy principle by the stopping criterion

$$\|G'(x_n)s - b_n^\delta\|_Y \leqslant \mu_n \|b_n^\delta\|_Y. \tag{8.5}$$

In our case the linearization error is of order $\mathcal{O}(\|x_n - x\|^2)$, so we may start with very crude solutions of the linearized equation, and then slowly decrease μ_n as we (hopefully) approach x. The method REGINN ("REGularization based on INexact Newton iteration"; see [LR06; Rie99; Rie01; Rie05]) provides a theoretical framework for this approach. It proves that there exists a choice of tolerances $(\mu_n)_{n\in\mathbb{N}}$ such that this whole process actually yields a regularization method, provided that the first x_n, that satisfies (8.3) is accepted as the solution. For the details we refer to the aforementioned articles. The most important insights (regarding the implementation) are that we must not let the tolerances converge to 0 or to 1 during the iteration. We must also take care that they do not become too small near the end of the iteration to prevent "over-satisfying" the discrepancy principle in the last REGINN-step.

CHOOSING THE TOLERANCES Since we have to give values for the tolerances μ_n ($n \in \mathbb{N}_0$), this method is also not free of hyper-parameters.

However, in practice the following strategy (which is based on the strategies presented in [Rie03] and [Win16]) to choose them seems to work quite well. It is based on the idea that we would like to *gradually* increase the difficulty of the linearized problems. A natural measure for this is the amount of steps k_i $(i = 0, \dots, n-1)$ that the inner regularization method needed to solve the previous problems. Given constants $\mu_{\text{start}}, \mu_{\max}, \zeta \in (0,1)$ and $\beta > 0$, we first calculate $\tilde{\mu}_n \in (0,1)$ by

$$\tilde{\mu}_n := \begin{cases} \mu_{\text{start}} & \text{if } n \leqslant 1, \\ 1 - \left(\frac{k_{n-2}}{k_{n-1}}\right)^{\beta} (1 - \mu_{n-1}) & \text{if } n \geqslant 2 \text{ and } k_{n-1} \geqslant k_{n-2} > 5, \\ \zeta \mu_{n-1} & \text{otherwise.} \end{cases} \tag{8.6}$$

A more conservative strategy (due to WINKLER [Win16]) is obtained by setting $\tilde{\mu}_0 := 1$, effectively accepting anything in the first step, followed by $\tilde{\mu}_1 := \|G(x_1) - y^{\varepsilon}\| / \|G(x_0) - y^{\varepsilon}\|$. In experiments, this resulted in slightly more computational effort than a well-tuned μ_{start}, but by eliminating μ_{start} it also avoids the risk of choosing it badly.

We use $\tilde{\mu}_n$ to define

$$\mu_n := \min\left\{ \mu_{\max},\ \max\left\{ \tilde{\mu}_n,\ \frac{\tau\varepsilon \|y^{\varepsilon}\|_Y}{\|G(x_{n-1}) - y^{\varepsilon}\|_Y} \right\} \right\}, \tag{8.7}$$

which ensures that the μ_n do not converge to 1. Further, it tries to safeguard the iteration from over-satisfying the discrepancy principle. The fraction in (8.7) is simply the ratio between the target discrepancy and the discrepancy that was already achieved in the previous REGINN-step. Therefore, μ_n is guaranteed to tend to $\mu_{\max}$ during the last steps of the algorithm.

Compared to [Rie03] and [Win16], we made two adjustments in this strategy: First, we do not regard inner step amounts smaller than five as enough computational effort to increase the tolerances. Otherwise, the tolerances tend to oscillate a lot in the early steps: The ratio k_{n-2}/k_{n-1} becomes small, even if the increase was just one step. Second, we do not multiply with $\mu_{\max}$ in every step; instead, we build minimums with it.

We make use of the parameters

$$\mu_{\max} := 0.999, \quad \mu_{\text{start}} := 0.9, \quad \beta := 1 \quad \text{and} \quad \zeta := 0.95$$

in all of the experiments to come.

SOLVING THE LINEARIZED PROBLEMS We still have to decide on a regularization method to use for the solution of the linearized problem (8.4) until the stopping criterion (8.5) is satisfied. Let $A_n := G(x_n)'$, then the problem reads $A_n s = b_n^{\delta}$. Originally, the theory for REGINN allowed only linear methods (that can be represented using filter functions), but it has been shown in [Rie05] that we may also use the conjugate gradient (CG) regularization method for this purpose, and we will do just that. Note that this regularization scheme is the application of the traditional

CG algorithm to the normal equation $A_n^* A_n s = A_n^* b_n^\delta$. Nevertheless, by exploiting the shape of the normal equations the algorithm becomes more stable and does not involve the calculation of inner products anymore. To be precise, it generates a sequence of iterates $(s_k)_k$ according to the following pseudocode.

Algorithm 8.1 (Conjugate Gradients on the Normal Equations)**.**
Given $s_0 \in X$, compute $(s_k)_k$ by

$r_0 \leftarrow b_n^\delta - A_n s_0, \;\; p^1 \leftarrow d^0 \leftarrow A_n^* r^0, \;\; k \leftarrow 1$
while $d^{k-1} \neq 0$ **do**
  $q_k \leftarrow A_n p_k$
  $\alpha_k \leftarrow \|d_{k-1}\|_X^2 / \|q_k\|_Y^2$
  $s_k \leftarrow s_{k-1} + \alpha_k p_k$
  $r_k \leftarrow r_{k-1} - \alpha_k q_k$
  $d_k \leftarrow A_n^* r_k$
  $\beta_m \leftarrow \|d^m\|_X^2 / \|d^{m-1}\|_X^2$
  $p_{k+1} \leftarrow d_k + \beta_k p_k$
  $k \leftarrow k + 1$
end while

If X is infinite-dimensional, then this sequence is usually infinite. Otherwise, the algorithm will terminate after a finite number of steps with the exact solution. However, as is common for iterative methods, the reciprocal of the number of CG-iterations assumes the role of the regularization parameter. This means we stop as soon as the first iterate satisfies the stopping criterion (8.5), i.e. at

$$k_n^* \coloneqq \min\{\, k \in \mathbb{N} \mid \|G'(x_n)s_k - b_n^\delta\| \leqslant \mu_n \|b_n^\delta\| \,\}.$$

Afterwards, we may set $x_{n+1} \coloneqq x_n + s_{k_n^*}$.

LINEAR SAFEGUARDING For the nonlinear problem we have the safeguarding rule (8.7). The same problem of possibly over-achieving the desired discrepancy can also occur in the linear problems, in particular for a fast Krylov-space based solver like CG. Since we are in a Hilbert space setting, we can interpolate between $s_{k_n^*}$ and its predecessor $s_{k_n^*-1}$ such that the target discrepancy is achieved exactly: We would like to find $\lambda \in [0, 1]$ such that

$$\begin{aligned} \mu_n^2 \|b_n^\delta\|^2 &= \|G(x_n)'[\lambda s_{k_n^*} + (1-\lambda) s_{k_n^*-1}] - b_n^\delta\|^2 \\ &= \lambda^2 \|G(x_n)' s_{k_n^*} - b_n^\delta\|^2 + (1-\lambda)^2 \|G(x_n)' s_{k_n^*-1} - b_n^\delta\|^2 \\ &\quad + 2\lambda(1-\lambda)\big(G(x_n)' s_{k_n^*} - b_n^\delta, G(x_n)' s_{k_n^*-1} - b_n^\delta\big). \end{aligned}$$

This is a quadratic equation in λ, which only involves known quantities: discrepancies of the last two CG-iterates, and the last two residuals. All of these are known during the iteration (for example the difference between old- and new residual is αq_k), thus computing λ results only in the evaluation of one inner product and can therefore be considered cheap.

THRESHOLDS FOR THE ITERATION COUNT Sometimes, the strategy (8.6) is not enough to ensure that the iteration numbers of CG increase gradually, because we might accidentally encounter very difficult linear subproblems. The conjugate gradient method provides a monotonous increase of its iterates' norms, thus in these cases it is of no harm if we simply stop the inner iteration and hope that this small step is enough to get back on track. Setting a constant maximum iteration count for CG to handle this problem is ill-advised, because we actually want them to increase. Instead, we take up the idea of [Win16] and set

$$k_n := \min\{k_n^*,\ k_n^{\max}\}, \qquad k_n^{\max} := \begin{cases} n+1 & \text{if } n \leqslant 1, \\ k_{n-1} + k_{n-2} & \text{if } n > 1, \end{cases}$$

which means that the current iteration may at most take as long as the last two iterations combined. As desired, they are still allowed to increase exponentially and in the worst case, the $(k_n)_{n\in\mathbb{N}_0}$ are equal to the Fibonacci sequence.

8.1.3 *Admissibility of iterates*

We have one last problem to deal with. As is usual for Newton-methods, the theory for REGINN states that the sequence $(x_n)_{n\in\mathbb{N}_0}$ is well-defined, i.e. belongs to $D(G)$, and is convergent to the true solution x if the initial guess $x_0 \in D(G)$ is *close enough* (with respect to the X-norm) to x. In practice, we cannot know a-priori whether this is the case, even if we know x. Furthermore, we have equipped all of our forward operators with L^2-based norms instead of L^∞. The pointwise constraints in $D(F_\rho)$ and $D(F_c)$ (as defined in equations (8.1) and (8.2)) only yield open subsets of $L^\infty(I; L^\infty)$. Since we have restricted our forward operators to a finite-dimensional space E, both norms are equivalent and $D(F_\rho) \cap E$ is open with respect to $L^2(I; L^2)$. However, the constant C such that $\|\cdot\|_{L^\infty} \leqslant C\|\cdot\|_{L^\infty}$ holds on E is likely going to be huge. For example, the basis functions of our finite element space V_h have a $L^\infty(\Omega)$-norm of 1 and a very small $L^2(\Omega)$-norm. This will further decrease the neighborhood around x in which we are allowed to start. Admissibility of the iterates is not only needed in the theoretical setup — a lack thereof will cause the linear systems in our wave equation solver to become singular. Hence, when $G \in \{F_\rho, F_c\}$ we will have to find a way that ensures that the REGINN-iterates x_n stay inside $D(F_\rho)$ and $D(F_c)$, respectively. Two ideas come to mind, which we will present on F_ρ. Without loss of generality we assume that $\rho_b = 0$.

PROJECTION Fix some $\delta^* > 0$. The subset

$$Z := \left\{ \rho \in W^{3,\infty}(I; L^\infty(\Omega)) \,\middle|\, \rho_0 + \delta^* \leqslant \rho \leqslant \rho_1 - \delta^* \;\text{ a.e. in } I \times \Omega \right\}$$

of $D(F_\rho) \cap W^{3,\infty}(I;L^\infty)$ is closed and convex, thus we can insert a projection onto it after each REGINN-step by defining

$$x_{n+1} \coloneqq P_Z(x_n + s_{k_n}).$$

This modification is quickly implemented, because we use piecewise linear elements in V_h, which (in contrast to quadratic elements) attain their maximal- and minimal values at mesh nodes. Thus, we can simply loop through the corresponding vectors and set values of ρ that are outside of $[\rho_0 + \delta^*, \rho_1 - \delta^*]$ to ρ_0 and ρ_1, respectively. However, this approach might cause the REGINN-iteration to not converge anymore.

TRANSFORMATION Another way to eliminate the constraints is to find a Fréchet-differentiable transformation

$$\Gamma\colon W^{3,\infty}(I;L^\infty(\Omega)) \to D(F_\rho) \cap W^{3,\infty}(I;L^\infty(\Omega))$$

and use $F_\rho \circ \Gamma$ as the forward operator instead of F_ρ, thereby reconstructing $\Gamma(\rho)$. Afterwards, we may obtain ρ by applying Γ^{-1} once. We will only consider transformations that operate pointwise, i.e. $(\Gamma f)(t,x) \coloneqq \gamma(f(t,x))$ for almost all $(t,x) \in I \times \Omega$ and $f \in W^{3,\infty}(I;L^\infty)$ with $\gamma\colon \mathbb{R} \to (\rho_0, \rho_1)$. Unsurprisingly, the properties of γ directly translate to Γ:

Lemma 8.2. *Assume that γ is bijective and four times continuously differentiable. Then Γ is a Fréchet-differentiable homeomorphism, and its derivative $\Gamma\colon W^{3,\infty}(I;L^\infty) \to \mathcal{L}\big(W^{3,\infty}(I;L^\infty)\big)$ is given by*

$$(\Gamma'(f)[h])(t,x) = h(t,x)\,\gamma'(f(t,x))$$

for almost all $(t,x) \in I \times \Omega$ and $f, h \in W^{3,\infty}(I;L^\infty(\Omega))$.

Proof. Note that γ already is a homeomorphism, and that Γ is well-defined: $\gamma\big([-\|f\|, \|f\|]\big)$ (with respect to $L^\infty(I;L^\infty)$) is a closed subinterval of (ρ_0, ρ_1) and therefore must have a positive distance to ρ_0 and ρ_1. Furthermore, Γ is also a homeomorphism. For the differentiability assertion we have to estimate

$$\begin{aligned}&\big(\Gamma(f+h)\big)(t,x) - \big(\Gamma(f)\big)(t,x) - \big(\Gamma'(f)[h]\big)(t,x)\\ &\quad = \gamma\big(f(t,x) + h(t,x)\big) - \gamma\big(f(t,x)\big) - h(t,x)\,\gamma'\big(f(t,x)\big)\end{aligned}$$

in $W^{3,\infty}(I;L^\infty(\Omega))$. This linearization error (and its derivatives) belong to $o(\|h\|)$ for almost all $(t,x) \in I \times \Omega$. The continuity of the derivatives of γ ensure that this convergence is uniform in (t,x). □

If we only had to enforce a lower bound, then the natural transformation to use would be $\gamma(s) = \rho_0 + \exp(s)$. Instead, we define

$$\gamma(s) \coloneqq \frac{\rho_0 + \rho_1}{2} + \frac{\rho_1 - \rho_0}{2}\tanh(s) \tag{8.8}$$

for $s \in \mathbb{R}$. We immediately see that this function is smooth and bijective, and that its first derivative is equal to

$$\gamma'(s) = \frac{\rho_1 - \rho_0}{2s^2}.$$

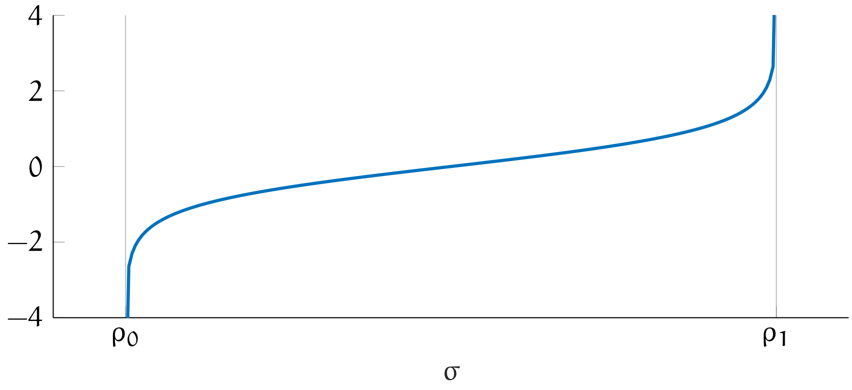


Figure 8.1: Inverse transformation γ^{-1} as given through (8.9)

Furthermore, its inverse at $t \in (\rho_0, \rho_1)$ is given as

$$\gamma^{-1}(\sigma) = \tanh^{-1}\left(\frac{2\sigma - \rho_0 - \rho_1}{\rho_1 - \rho_0}\right). \tag{8.9}$$

We have included a graphical representation of γ^{-1} as Figure 8.1. Note that for $\rho_0 \ll \sigma \ll \rho_1$, the function $\gamma^{-1}(\sigma)$ is almost linear, hence we only perturb the parameter at points where its values are close to ρ_0 or ρ_1.

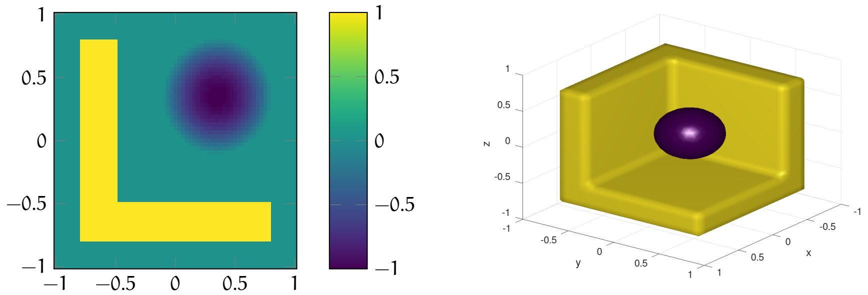
In numerical tests we made the observation that projecting the estimates onto the constraints performs well for high noise levels $\varepsilon \gtrapprox 1\%$, but as suspected it causes divergence for small ε. Thus we will make use of the transformation approach in all experiments for c and ρ.

8.2 INVERSION SCENARIOS

We will now present two functions that we want to use to build exact parameters for each of the inverse problems. To ease the presentation, we consider $d \in \{2, 3\}$ and $\Omega := [-1, 1]^d$ in the sequel. The time interval is given by $I = [0, T]$ with $T = 2\pi$.

We know from our theoretical results that we may use every essentially bounded function for q and ν, while for c and ρ it must be smooth in time ($W^{1,\infty}$ to make F well-defined and $W^{3,\infty}$ to make it Fréchet-differentiable), in addition to being bounded away from zero. We obtain such a function by smoothly scaling a space-dependent function. In space, we want it to consist of discontinuities and smooth parts.

For the discontinuous part we make use of a piecewise constant function. For this, let $\Omega_L(\alpha, \beta) \subset \Omega$ denote an L-shape with distance $\alpha > 0$ from the coordinate axes and width $\beta > 0$. While for $d = 2$ this description should be sufficient, it is not clear what it means in three dimensions. In an attempt to make the d-dimensional volume of these sets approxi-



(a) $d = 2$

(b) $d = 3$

Figure 8.2: Test parameter Λ_{LDot}, evaluated at $t = \pi/2$

mately equal, we define it to be the union of three thin plates, each aligned with one coordinate plane. In both cases, this set can be described by

$$\Omega_L(\alpha, \beta) = \Big\{ x \in \Omega \;\Big|\; \text{there is } i_0 \in \{1, \dots, d\} \text{ s.t. } x_{i_0} \in [\alpha - 1, \alpha + \beta - 1] \\ \text{and } x_i \in [\alpha - 1, 1 - \alpha] \text{ for } i \neq i_0 \Big\}.$$

For the continuous part we define

$$\lambda_{\omega,r}(x) := \begin{cases} 1 - r^{-2} \|x - (\omega, \dots, \omega)\|^2 & \text{if } \|x - (\omega, \dots, \omega)\| \leqslant r, \\ 0 & \text{otherwise.} \end{cases}$$

We combine Ω_L and λ by defining our first test parameter to be

$$\Lambda_{\mathrm{LDot}}(t, x) := \left(0.2 + 0.8 \cdot \sin(t)^2\right) \left(\mathbb{1}_{\Omega_L(0.2,0.3)}(x) - \lambda_{0.35,0.45}(x)\right).$$

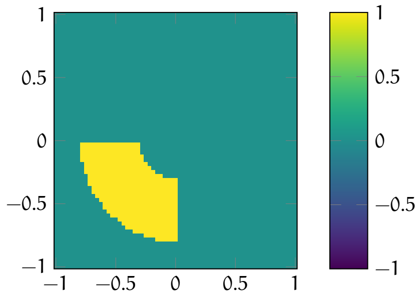
Its discretization is depicted in Figure 8.2. The meshes we employed for this task are the same meshes that we will be using for the inverse problems. They consist of $N = 256$ discretization points in time and $M = 4225$ (if $d = 2$) or $M = 35937$ (if $d = 3$) nodes in the finite element space V_h.

In order to test how sharp these smoothness constraints are, we will also try to reconstruct another function that specifically does not satisfy them. Again, we work with characteristic functions, but instead of scaling them we make the underlying set time-dependent. For $d = 3$, this set is given as the spherical shell segment

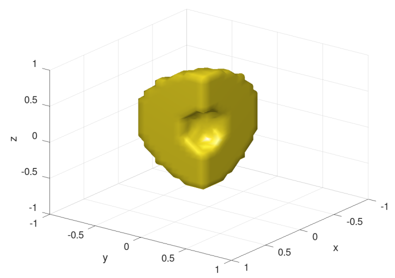
$$\Omega_R(r_1, r_2, t) = \Big\{ (r \sin\theta \cos\varphi, r \sin\theta \sin\varphi, r \cos\theta) \in \Omega \;\Big|\; r \in [r_1, r_2], \\ \theta \in [0, \pi] \text{ and } (\varphi - t) \bmod 2\pi \in [0, \pi/2] \Big\},$$

and we extend this definition to $d = 2$ by regarding $\mathbb{R}^2$ as a subset of $\mathbb{R}^3$ using the embedding $(x, y) \mapsto (x, y, 0)$. Note that the characteristic function

$$\Lambda_{\mathrm{Ring}}(t, x) := \mathbb{1}_{\Omega_R(0.3,0.8,t)}(x)$$



(a) $d = 2$



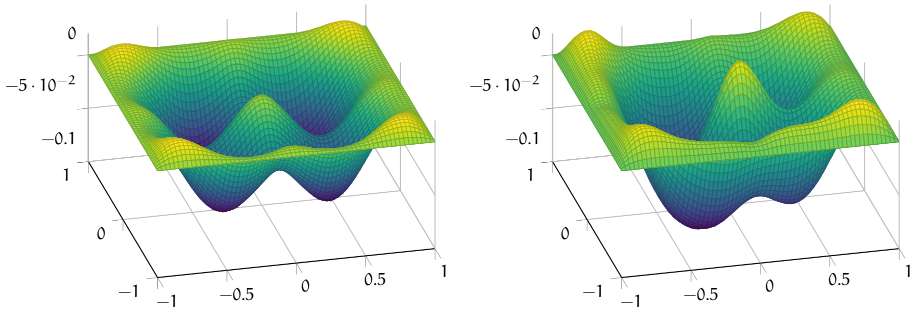
(b) $d = 3$

Figure 8.3: Test parameter Λ_{Ring}, evaluated at $t = \pi$

is discontinuous in x and, more importantly, also in t. It is shown in Figure 8.3.

Naturally, both parameters become continuous in space after discretization, because our finite element space V_h consist of continuous functions, and a similar argument holds for the time variable. However, the corresponding slopes tend to be very large (especially if we refine even more). It will be interesting to see whether this causes any problems in the reconstruction process.

BACKGROUND VALUES AND SCALING As explained in Section 8.1, our parameters will be composed of a known background and a "small" perturbation. For problems concerning q and ν, we will set the background functions q_b, ν_b to zero and for ρ and c we want to use a positive constant which we have to choose properly. The function c_b determines how fast the waves in Ω propagate. If we set it too high, then any movement induced by the right-hand side f will almost instantaneously propagate through the whole domain. This makes for good data for the inverse problem, but it seems unrealistic. However, there is also the risk of choosing it too low. In this case, a localized f will result in a wave-front that slowly travels through the domain Ω and will therefore only illuminate a small part of it. This means that the data u cannot contain much information about the unknown parameter, as discussed in Section 5.2.1. We would like to find a compromise between these two extremes: Wave propagation should be visible but the wave should also be able to reach the whole domain in time T, even if the right-hand side is very localized in space. Experiments show that for our domain this is the case for $c_b(t, x) := 0.3$. The scaling of ρ_b is irrelevant if $q = \nu = 0$, because then $1/\rho$ appears in every term of the left-hand side of the acoustic wave equation (5.15). Thus, we will use $\rho_b(t, x) := 1$. We need to fix constants for the transformation function γ (as defined in (8.8)) that restrict the values of the reconstructed ρ and c so that $\rho + \rho_b$ and $c + c_b$, as well as their reciprocal values, are uniformly positive. We merely



(a) $F_c(0)$ at $t \approx 3.45$ (b) $F_c(0.15\Lambda_{\text{LDot}})$ at $t \approx 3.45$

Figure 8.4: Data of the initial guess compared to the exact data for $c = 0.15\Lambda_{\text{LDot}}$

wish to avoid running into numerical difficulties, hence we choose big intervals. Precisely, they are given by

$$c(t,x) \in (c_0, c_1) := (-0.27, 30) \quad \text{and} \quad \rho(t,x) \in (\rho_0, \rho_1) =: (-0.9, 100).$$

In applications, one might have a better idea about the magnitude of the unknown parameters. This can then be incorporated in γ as a-priori knowledge.

Since we do not have a specific application in mind it is difficult to know how we should scale Λ_{LDot} or Λ_{Ring} in order to obtain "realistic" perturbations. For example, if we end up using a very large q, then the term qu will dominate the left-hand side of the acoustic wave equation, and the inverse problem degenerates to finding a stable approximation for $q \approx f/u$. The other extreme happens if we scale the perturbation too small, because then our regularization method will find that the initial guess already satisfies the discrepancy principle. We chose the scaling coefficient α heuristically, and in such a way that there is a visible difference in the data compared to the data with only the background parameters. This approach resulted in using $\alpha = 0.15$ for F_c, $\alpha = 10$ for F_ν, $\alpha = 0.75$ for F_ρ and $\alpha = 20$ for F_q. Figure 8.4 contains snapshots of $F_c(0) = F(c_b, \nu_b, \rho_b, q_b)$ and $F_c(0.15\Lambda_{\text{LDot}})$ at $t \approx 3.45$. As desired, we can clearly see that the perturbed parameter creates differences in the data, but these differences do not seem to fundamentally change the wave propagation.

8.3 CONVERGENCE

Let $G\colon E \to L^2(I; L^2)$ again denote one of our forward operators F_c, F_ν, F_ρ and F_q, possibly transformed by Γ. One of the most important aspects of any regularization method is its behavior when the noise level ε in the data tends to zero. In particular, the regularized solution x^ε of $Gx = y^\varepsilon$ should converge to the ground truth x. For REGINN this is the case if the initial guess x_0 is close enough to x and if one chooses the tolerances correctly. However, in order for these theoretical results to be applicable, we have to assume that G satisfies the nonlinearity condition. We were only able to prove this for the (not so interesting) cases $G \in \{F_q, F_\nu\}$

in Section 4.2.1. The convergence rates are then determined by source conditions for the *exact parameter* x. In practice, one also typically does not have knowledge about the size of the neighborhood around x in which we are supposed to start. Hence, we actually do not know at all if our implementation of CG-REGINN will be able to reliably invert any of our forward operators. In this section, we will try this for the two example parameters $x = \alpha\Lambda_{\text{LDot}}$, $x = \alpha\Lambda_{\text{Ring}}$ that we presented (along with the scaling factors α) in the previous section.

In order to converge to the ground truth and not another parameter, we must try to ensure injectivity of the forward operator. Therefore we will not use a localized right-hand side here; instead we set $f(t,x) := 2\cos t$. Also note that we use the discretized parameters to generate the data, so that the exact data $u = Gx$ is guaranteed to belong to the range of the discretized forward operator (which also implies that we commit the "inverse crime"). After the data generation we set

$$u^{\varepsilon} := u + \varepsilon \|u\|_{L^2(I;L^2(\Omega))} \frac{\eta}{\|\eta\|_{L^2(I;L^2(\Omega))}}, \tag{8.10}$$

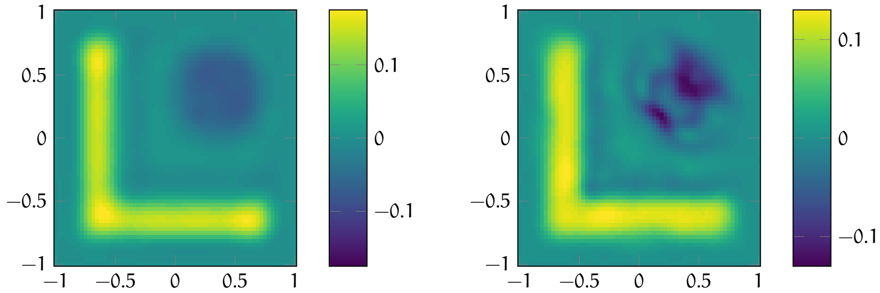
where η contains white noise that is uniformly distributed in $[-1,1]$.[1]

For REGINN we make use of the tolerance strategy and other modifications as discussed in Section 8.1, with initial guess $x_0 = 0$ and $\tau = 2$. Naturally, we use the $L^2(I;L^2(\Omega))$-norm for the image space of G, but we have multiple options for its preimage space E. In Section 7.2, we have derived approximations for the L^2-, H^1- and $H^2(I;L^2(\Omega))$ norms. More importantly, we are able to transform $G'(x)^*$ to conform to these norms. Thus, we will reconstruct the parameter with one of these. However, we would like to avoid accidentally shifting the focus too far away from the L^2-part. To this end, we set the weights in the H^1- and $H^2(I;L^2(\Omega))$-norms to $1/2$ in front of the first-order derivatives, and $1/4$ in front of the second-order term in the H^2-norm.

In the two-dimensional setting we used eleven different values for ε, logarithmically spaced between 10^{-4} and 0.22 for each of these 24 configurations (4 forward operators, 3 different norms and 2 test parameters). For $d = 3$ we used eight different noise levels between 10^{-3} and 0.22, and only performed the tests for the $H^1(I;L^2(\Omega))$-norm in order to keep the computing times manageable. All in all, this involved the solution of approximately 1.5 million partial differential equations in about three years of CPU time (on four eight-core desktop computers and one twenty-four-core computing server).

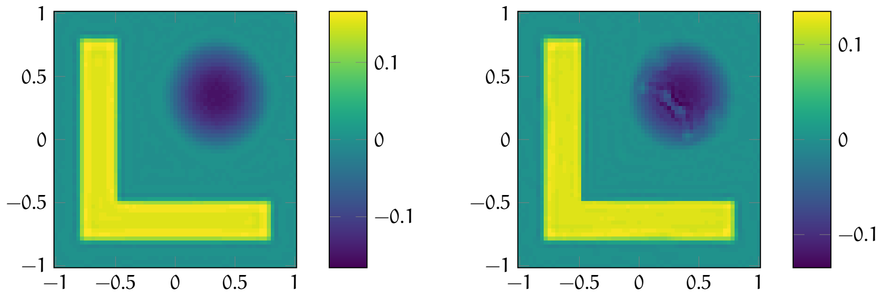
Obviously we cannot present all of the obtained results. Instead, we will consider the reconstruction of Λ_{LDot} with the $H^1(I;L^2(\Omega))$-norm as the "base" case and analyze it thoroughly. Then we will vary the ground truth and the norm of the preimage space, and only remark on what changed by doing so. Tables containing all of the measured convergence rates can be found in Appendix B.

1 This noise is obtained by the default pseudorandom number generator in the C++ standard library after seeding it with the current time.



(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.5: Reconstruction of $c = 0.15\Lambda_{\mathrm{LDot}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-2}$



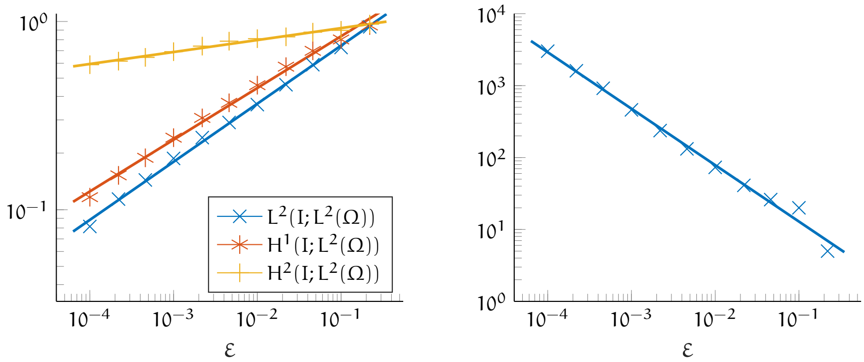
(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.6: Reconstruction of $c = 0.15\Lambda_{\mathrm{LDot}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$

8.3.1 *Smooth parameter in* $H^1(I; L^2(\Omega))$

We start with the two-dimensional setting and Λ_{LDot} as the ground truth. If we equip the preimage space E of F_c with the $H^1(I; L^2(\Omega))$-norm, then we obtain a reconstruction as shown in Figure 8.5. There the noise level was set to one percent. For the presentation we chose to evaluate the reconstructed parameter at two time instances: $t = \frac{1}{2}\pi$ is near the start of the simulation, and $t = \frac{4}{3}\pi$ is closer to $T = 2\pi$. Although both images are blurry, we can clearly recognize the resemblance to the ground truth (cf. Figure 8.2a). We also note that there is a small "shadow" around the L-shape in both pictures, and that the approximation near the ball is better for the lower value of t. If we go to a very low noise level of $\varepsilon = 10^{-4}$, as seen in Figure 8.6, then this shadow is gone. It is noteworthy how well the L-shape is approximated; in its vicinity the reconstruction seems to be piecewise constant, although we are using L^2-based norms that do not enforce sparsity of ∇c at all (as opposed to TV-regularization). However, we should also note that the edges of this shape are perfectly aligned with the mesh, which can be seen as a-priori information. Again, the approximation of the ball-shaped part of Λ_{LDot} seems to be better for $t = \frac{1}{2}\pi$, because for $t = \frac{4}{3}\pi$ some artifacts begin to develop.

The relative $H^1(I; L^2)$-errors for $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-4}$ are 46 % and 12 %, respectively. These values are also depicted in Figure 8.7a, along with the errors for nine other values for ε. In the figure we have also included the corresponding $L^2(I; L^2)$- and $H^2(I; L^2)$ errors. Regardless of



(a) Reconstruction errors (b) Calls to PDE solver

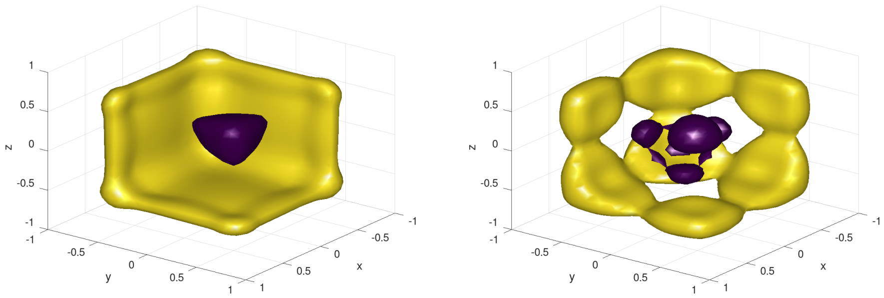
Figure 8.7: Behavior of the errors and computing times for the reconstruction of 2D $c = 0.15\Lambda_{\text{LDot}}$ using $H^1(I; L^2(\Omega))$ depending on the noise level ε

the norm we use to gauge the error, on a logarithmic scale they describe straight lines. If we assume this behavior to hold for $\varepsilon \to 0$, then we are able to conclude the rates

$$\|c - c^\varepsilon\|_{L^2(I;L^2)} \approx \mathcal{O}\big(\varepsilon^{0.31}\big), \ \|c - c^\varepsilon\|_{H^1(I;L^2)} \approx \mathcal{O}\big(\varepsilon^{0.27}\big), \\ \text{and} \ \ \|c - c^\varepsilon\|_{H^2(I;L^2)} \approx \mathcal{O}\big(\varepsilon^{0.06}\big).$$

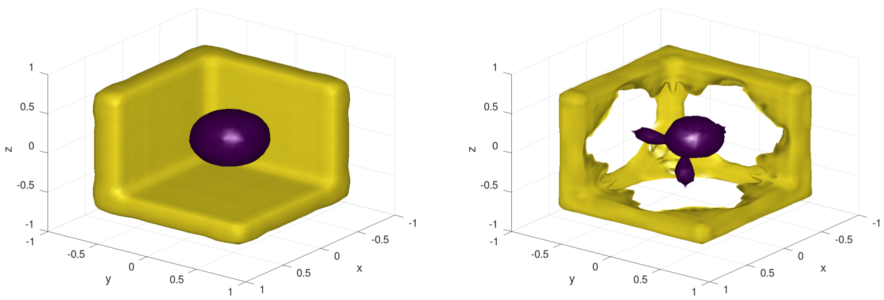
The slope for the H^2-error is so small that we cannot really speak of "convergence". The error seems to be almost unaffected by the noise level, although it could theoretically even increase. The most important rate is the one for the H^1-norm, because this is the norm we used to obtain c^ε. Convergence rates provided by the theory for REGINN (under some assumptions like the tangential cone condition, cf. [Rie05]) are of the form $\|c - c^\varepsilon\| = \mathcal{O}\big(\varepsilon^{(\kappa - C)/(1+\kappa)}\big)$ for a constant $C > 0$, if the exact parameter satisfies the source condition $c = |F_c'(c)|^\kappa w$ for some $w \in E$ and $\kappa \in (C, 1]$. Hence, the highest guaranteed rate of convergence (if $\kappa = 1$ and C is negligible) is $1/2$. Pretending that these theoretical results hold for our operator, we are led to conclude that $c = 0.15\Lambda_{\text{LDot}}$ fulfills such a source condition for some $\kappa > 0.38$. The fact that the convergence rate in the $L^2(I; L^2)$-norm is only slightly higher implies that the H^1-error is not dominated by $\big\|c' - (c^\varepsilon)'\big\|_{L^2(I;L^2)}$. These rates are also prototypical for the other forward operators, in particular for ρ. We would like to remind the reader that all convergence rates can be found in Appendix B.

COMPUTATIONAL EFFORT In practice, it is also interesting to see how the effort in obtaining the reconstruction c^ε increases as $\varepsilon \to 0$. Obviously, the time that is needed to run REGINN until the discrepancy principle is satisfied is highly dependent on the system and the implementation. Instead, we measure the effort by how many solutions of the acoustic wave equation were needed in the whole reconstruction process. Note that each evaluation of F_c, F_c' and $(F_c')^*$ requires the solution of such a partial differential equation. Figure 8.7b shows the behavior of this quantity for the different noise levels. From it we deduce that this effort



(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.8: Reconstruction of $\rho = 0.75\Lambda_{\mathrm{LDot}}$ in $H^1(I; L^2(\Omega))$ in 3D with $\varepsilon = 4.6 \cdot 10^{-2}$

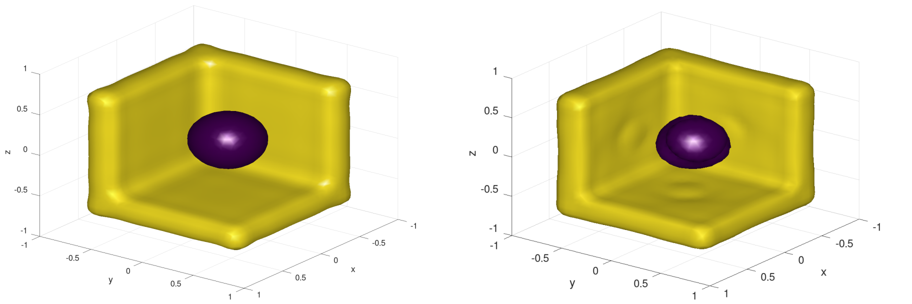


(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.9: Reconstruction of $\rho = 0.75\Lambda_{\mathrm{LDot}}$ in $H^1(I; L^2(\Omega))$ in 3D with $\varepsilon = 10^{-3}$

is of order $\mathcal{O}(\varepsilon^{-0.8})$. For ρ, q and ν we have observed similar rates, ranging from $\varepsilon^{-0.7}$ to ε^{-1}. Unfortunately, it is difficult to compare them with theoretical predictions. On the theoretical side we can only obtain a bound for the number of REGINN-steps, but these involve a highly variable amount of inner CG-iterations and are therefore not useful to estimate the computational effort. However, it is noteworthy that these rates are on a par with the theoretical results for how the CG-method might perform to solve a *linear* inverse problem (see [EHN00]), so we can consider our results to be very encouraging.

THE THREE-DIMENSIONAL CASE An example reconstruction of $\rho = 0.75\Lambda_{\mathrm{LDot}}$ for $d = 3$ and a quite high noise level of 4.6% (which means that we stop at 9.2% of discrepancy) is shown in Figure 8.8. To be more precise, the figure shows iso-surfaces at half of the maximal- and half of the minimal value at the respective time step. Although the overall $H^1(I; L^2(\Omega))$-error amounts to 70 %, at least at $t = \frac{1}{2}\pi$ the support of the ground truth is already well approximated. If we go to 0.1% noise, seen in Figure 8.9, the edges of the L-shape become much sharper. We can also see the ball where the smooth part of the ground truth is located. However, as we get closer to $T = 2\pi$, the negative value of this smooth part still "bleeds out" into the L-shape. In this case, the total error decreased to about 40 %.



(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.10: Reconstruction of $q = 20\Lambda_{\mathrm{LDot}}$ in $H^1(I; L^2(\Omega))$ in 3D with $\varepsilon = 10^{-3}$

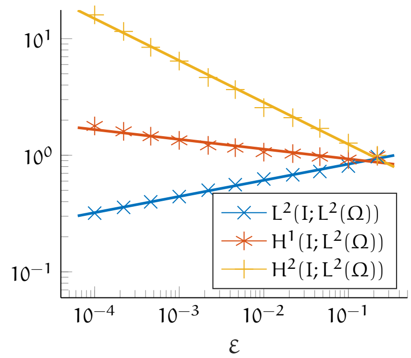
For the other three forward operators, the support of the 3D version of Λ_{LDot} is better reconstructed. For example, F_q admits a reconstruction as in Figure 8.10 when using 0.1% noise. The convergence rates are also slightly higher than for F_ρ. For F_c, we observed a divergence of REGINN for noise levels lower than 1%. In these cases, the discrepancy decreased to about 1%, only to increase again after that. The following iterates begin to develop singularities around the ball-shaped part. We have already observed a similar behavior in 2D (cf. Figure 8.6), but much less pronounced. This could indicate that the $H^1(I; L^2(\Omega))$-norm might not be enough to guarantee the required smoothness in time in every case.

8.3.2 *Influence of the reconstruction norm*

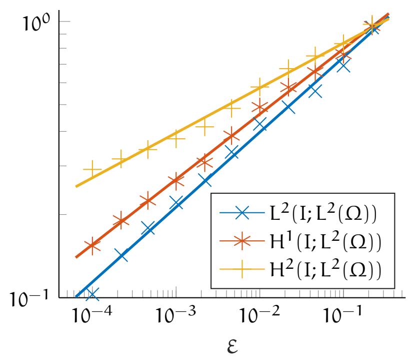
Apart from the $H^1(I; L^2(\Omega))$-norm, we have two other norms at our disposal which we can use for the preimage space E of the forward operator. From our theoretical analysis we know that the parameter q should be okay with only $L^\infty(I; L^\infty(\Omega))$. In [GL17] we have shown that in the spatial variable this can be relaxed to $L^2(\Omega)$. Based on these results we can be optimistic about the reconstruction of q using this norm. For ν, we know that the differentiability in time is only needed to ensure uniqueness of a weak solution of the PDE in the continuous context. Hence, $L^2(I; L^2(\Omega))$ might also be enough for the identification of the attenuation coefficient ν.

Indeed, for both q and ν, reconstructions using the $L^2(I; L^2(\Omega))$- or $H^2(I; L^2(\Omega))$-norm do not reveal any difficulties. As one would expect, the difference is not felt when viewing snapshots of the reconstruction at fixed time instances. However, by interpreting them as videos it becomes clear that the H^2-norm indeed forces the REGINN-iterates to vary much more smoothly in time. Convergence plots for F_q can be found in Figure 8.11.

Unsurprisingly, using the H^2-norm in time causes the H^2-error to decrease when $\varepsilon \to 0$, in this case with speed $\mathcal{O}(\varepsilon^{0.16})$. However, the errors do not describe a straight line; one can already observe some saturation that suggests a smaller slope for even smaller noise levels. Furthermore, the H^2-norm produces slightly higher L^2- and H^1- convergence rates than



(a) Using $L^2(I;L^2(\Omega))$



(b) Using $H^2(I;L^2(\Omega))$

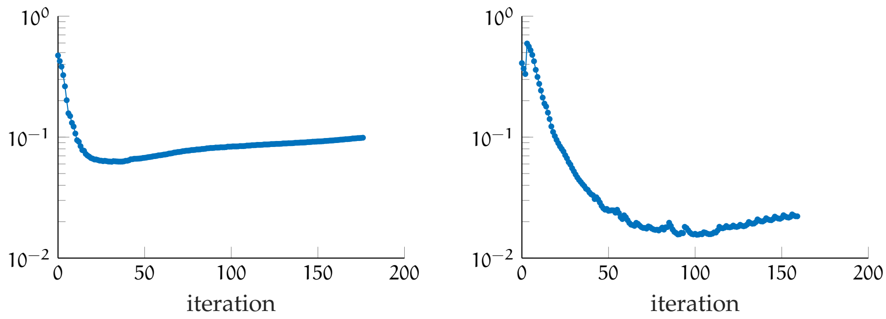
Figure 8.11: Reconstruction errors for $q = 20\Lambda_{\mathrm{LDot}}$ in 2D for different reconstruction norms

when using $H^1(I;L^2(\Omega))$. For example, the rate for the $H^1(I;L^2(\Omega))$-error increased from 0.21 to 0.24.

We have observed that using the H^1-norm already ensures that the H^2-error does not increase for decreasing noise level. This is not the case when using $L^2(I;L^2(\Omega))$. There, both the H^1- and H^2-error become unbounded in the limit $\varepsilon \to 0$; even the L^2-convergence rate drops. The asymptotic behavior of the required computational effort is approximately the same for all norms, but the constants are very different: Here, H^1-reconstructions are about 20% faster to obtain than those in L^2 or H^2. In conclusion, there seems to be no reason to prefer a reconstruction of q or ν in $L^2(I;L^2)$ over $H^1(I;L^2)$- or even $H^2(I;L^2)$. Naturally, here we knew a-priori that the ground truth is very smooth in time, which might not always be the case.

We are less optimistic for the combination of $L^2(I;L^2(\Omega))$ with F_c and F_ρ, because even to obtain a well-defined solution operator these operators already need to be weakly differentiable once. More than that, for the existence of the Fréchet-derivative they are required to be three times differentiable in time. Indeed, for most noise levels, the corresponding REGINN-iterations diverge. The behavior of the discrepancies in case of $\varepsilon = 10^{-4}$ are shown in Figure 8.12. Initially, both the discrepancy and the relative error to the ground truth decrease. However, at some point the iterates begin to develop sharp oscillations everywhere in the domain, which causes the discrepancy and the error to increase again. The lowest noise levels, for which the algorithm terminated successfully are 1% for F_ρ and 10% for F_c.

When reconstructing c or ρ with the $H^2(I;L^2(\Omega))$-norm, we would expect the same behavior as for q and ν, and no difficulties. For ρ this is the case. Surprisingly, the inversion of F_c fails for noise levels ($\varepsilon < 10^{-3}$), and shows a similar oscillatory behavior as when using the $L^2(I;L^2(\Omega))$-norm near the ball-shaped part of Λ_{LDot}. Since the noise levels where this happens are very small, we attribute this to numerical inaccuracies that effectively contribute to the noise. The right-hand side for the derivative of F with respect to c involves another finite difference scheme, which



(a) F_c (b) F_ρ

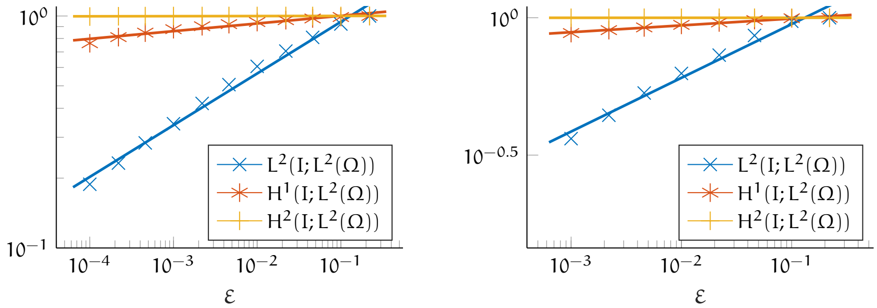
Figure 8.12: Relative discrepancies of REGINN-iterates during the reconstruction of Λ_{LDot} using $L^2(I; L^2(\Omega))$ and $\varepsilon = 10^{-4}$

contributes to the overall approximation error for the derivatives. Thus, it is not entirely unexpected that the parameter c is the first to show strange behavior for low noise levels.

8.3.3 *Violating the smoothness constraints*

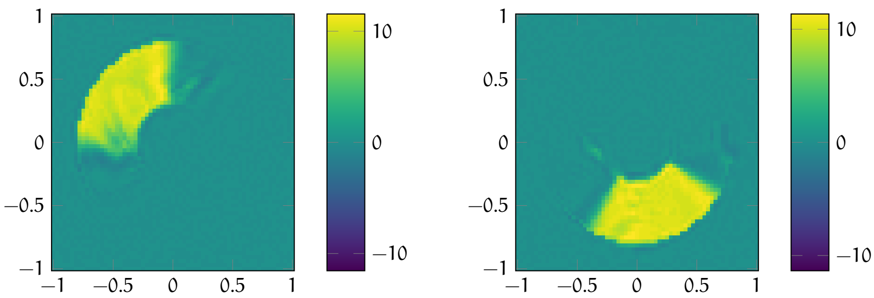
In Section 8.2 we have designed the parameter Λ_{Ring} specifically in order to be able to test nonsmooth parameters. This parameter is discontinuous in time. In principle, the remarks in the beginning of the previous subsection still apply; in particular, we expect this parameter to be reconstructed successfully if we use it for F_q or F_ν. Indeed, no matter which norm we use, we obtain $L^2(I; L^2(\Omega))$-convergence for these two forward operators. Because the ground truth itself does not belong to $H^1(I; L^2(\Omega))$, there is no convergence whatsoever for the $H^1(I; L^2(\Omega))$ or even $H^2(I; L^2(\Omega))$-error. Nevertheless, the best convergence speed in $L^2(I; L^2(\Omega))$ is again obtained by at least enforcing some regularity in time through the H^1-norm. For F_ν, the corresponding reconstruction errors are the subject of Figure 8.13, which further shows that this behavior seems to be independent of the space dimension. Example reconstructions (in 2D) for a low noise level are depicted in Figure 8.14. We see that the discontinuities at both the outer- and the inner radius of the ring are approximated well. This not the case at the two other, moving edges. Naturally, their blurring is a consequence of the chosen reconstruction norm. It vanishes if we instead reconstruct in $L^2(I; L^2(\Omega))$, as can be observed in Figure 8.15. However, there we have to deal with new artifacts *inside* the ring-shape.

For c and ρ all of the reconstruction attempts of Λ_{Ring} for low noise level diverge, no matter the reconstruction norm. This happens as soon as the inversion algorithm has to begin to approximate the discontinuities in order to further decrease the discrepancy, which usually does not happen for high noise levels. The mass density seems to be a bit more robust here, there we can go as low as a few per-mil of noise. The reconstruction of such a wave speed already fails for noise levels lower than one percent, because the discrepancy stops decreasing after reaching



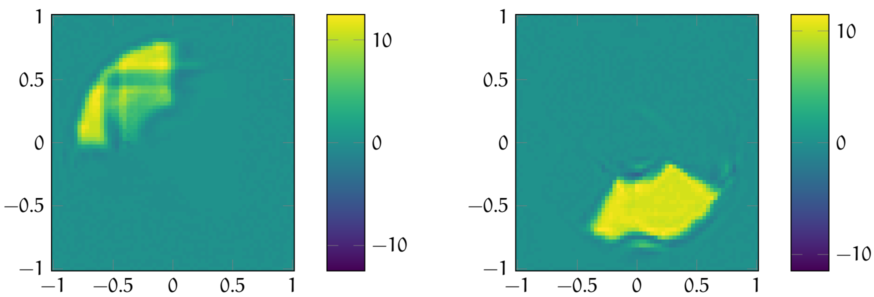
(a) two-dimensional setting (b) three-dimensional setting

Figure 8.13: Reconstruction errors for 2D $\nu = 10\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$



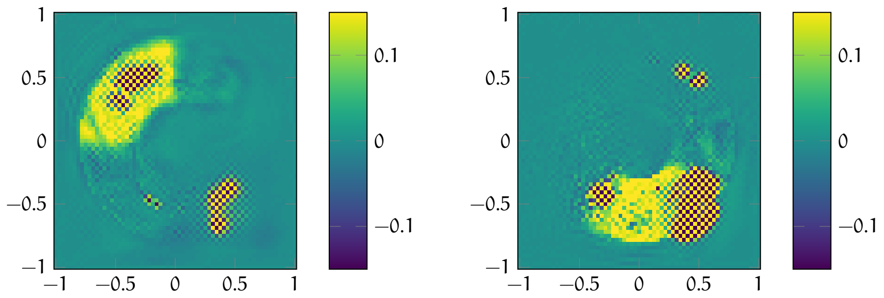
(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.14: Reconstruction of $\nu = 10\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$



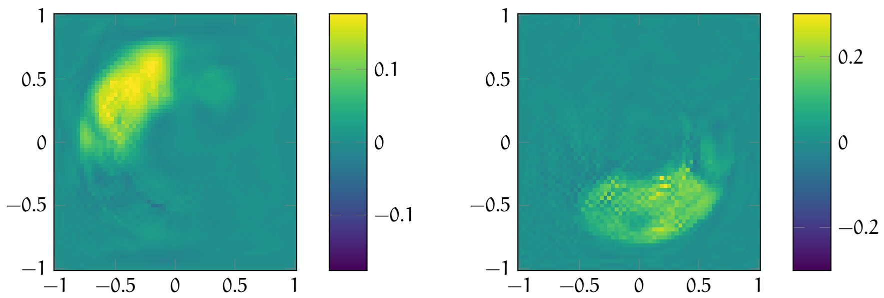
(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.15: Reconstruction of $\nu = 10\Lambda_{\text{Ring}}$ in $L^2(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$



(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.16: Last REGINN-iterate for $c = 0.15\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-4}$



(a) evaluated at $t = \frac{1}{2}\pi$ (b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.17: Reconstruction of $c = 0.15\Lambda_{\text{Ring}}$ in $H^1(I; L^2(\Omega))$ in 2D with $\varepsilon = 10^{-2}$

this threshold, no matter the noise level. After that, the iterates begin to develop singularities in the ring segment. A failed reconstruction for $\varepsilon = 10^{-4}$ (using the $H^1(I; L^2(\Omega))$-norm) can be found in Figure 8.16. Note that values exceeding $[-0.15, 0.15]$ are clipped, and the oscillations have values in $[-2, 2]$. However, we would like to remark that one percent of relative noise is already quite low with regards to applications, and the reconstruction for $\varepsilon = 10^{-2}$ already looks satisfying, see Figure 8.17.

8.4 Reconstruction from partial data

In this section, we will attempt the inversion of F_c and F_ρ in a more "realistic" measurement setup. In our convergence experiments, both the excitation of the wave inside Ω and its measurements were nonlocalized in space. We chose this setup to make certain that the data contains a lot of information about the ground truth. In applications, the setup will probably not be so convenient. For example, in the setting of non-destructive testing, we can think of a plate (e.g. made of carbon fiber reinforced polymers) that is under test, like in [LS17]. There, vibrations inside the plate are excited near localized points at the plate's boundary, by means of a Piezo-electric crystal or a laser beam. These vibrations are then picked up by other piezo crystals that are glued to the plate.

To conclude, we make two changes to our forward operators F_c and F_ρ to obtain a more realistic scenario: First of all, we replace the right-hand side $f(t,x) = \cos(2t)$ by localized pulses. Second, we compose it with a measurement operator. All in all, this will produce a setup similar to [GL17]. We briefly outline both changes, and then analyze what effects they have on the numerical reconstruction. For ease of presentation we assume that the operator under consideration is F_c. Furthermore, we only consider $d = 2$ here, because the three-dimensional case mainly differs in the computing times.

EXCITATION BY LOCALIZED PULSES To account for the resulting loss of data quality we assume that we have multiple measurements for different right-hand sides $f_1, \dots, f_K$ available. Of course we can model this by replacing $F_c \colon c \mapsto u$ by the vector-valued forward operator

$$\hat{\mathbf{F}}_c \colon E \to L^2(I; L^2(\Omega))^K, \quad \hat{\mathbf{F}}_c(c) = (u_1, \dots, u_K),$$

where u_k ($k = 1, \dots, K$) denotes the solution of the acoustic wave equation (5.15) for the right-hand side f_k. Since we only work with a finite amount $K \in \mathbb{N}$ of right-hand sides, we can apply all of our previous theoretical results on each component function to obtain the corresponding result for the whole $\hat{\mathbf{F}}_c$. The implementation of such a vector-valued forward operator is straightforward. It even offers a new, exciting possibility for parallelization: The process of solving the acoustic wave equation can be performed independently for each right-hand side, for example on different computers. Afterwards, these processes only have to distribute the computed wave fields to one another, e.g. using MPI (Message Passing Interface). This nicely complements the multi-threaded approach for matrix assembly (and the solution of linear systems) employed by the finite element library DEAL.II.

In the inverse problem we assume that $(u_1^\varepsilon, \dots, u_K^\varepsilon)$ is given. For us it is convenient to assume that the relative noise level ε is chosen with respect to the euclidean norm of the whole tuple, i.e.

$$\sum_{k=1}^{K} \|u_k^\varepsilon - u_k\|_{L^2(I;L^2)}^2 \leqslant \varepsilon^2 \sum_{k=1}^{K} \|u_k\|_{L^2(I;L^2)}^2.$$

Note that this setup implies that each field is allowed to carry a different noise level. However, in our numerical implementation we use a random-number generator to create additive white noise $(\eta_1, \dots, \eta_K)$ and then scale this whole tuple accordingly, as done in (8.10). Due to the high number of degrees of freedom for each η_k, it is very improbable that the norms of the η_k will exhibit huge variations. This also means that the η_k are pairwise uncorrelated, i.e. we expect (η_j, η_k) to be negligible. As a consequence, even in the case that all the right-hand sides f_k are identical, the data quality for $K > 1$ will by higher than for $K = 1$. Indeed, in this case we have $u_1 = \dots = u_K =: u$ and each field u_k^ε would carry a relative noise level of approximately ε. Since the noise is uncorrelated, we expect their average to have a noise level of $\varepsilon/\sqrt{K}$, as can be seen on

$$\left\| u - \frac{1}{K}\sum_{k=1}^{K} u_k^{\varepsilon} \right\|^2 \approx \frac{1}{K^2}\sum_{k=1}^{K} \|u - u_k^{\varepsilon}\|^2 \approx \frac{\varepsilon^2}{K}\|u\|^2.$$

We imagine that the domain $\Omega = [-1,1]^2$ can be excited near all of its edges, i.e. set $K \coloneqq 4$. Let $x_1, \dots, x_4$ denote the midpoints of these edges, counting clockwise starting from $x_1 \coloneqq (0,1) \in \partial\Omega$, and

$$f_k(t,x) \coloneqq \sin(2t)\,\mathbb{1}_{B(x_k,0.1)}(x)$$

for $(t,x) \in I \times \Omega$ and $k = 1, \dots, K$.

We would like to remark that this naïve approach of combining all measurements into a single forward operator can be hazardous: Even if the discrepancy principle is already satisfied for some of the fields, they can still be used by the inversion algorithm to decrease the overall discrepancy, potentially leading to more artifacts in the reconstruction. We did not run into such problems, maybe because the norms of the four data sets are almost equal. However, for inhomogeneous measurement data one should consider more specialized techniques, for example the Kaczmarz-variant of REGINN [MRL14].

SIMULATION OF MEASUREMENTS We restrict our analysis to *linear* measurement operators $\Psi\colon L^2(I; L^2(\Omega)) \to Z$ for some Hilbert space Z, which we compose with each component of $\hat{\mathbf{F}}_c$. This gives rise to

$$\mathbf{F}_c\colon E \to Z, \quad \mathbf{F}_c(c) \coloneqq (\Psi u_1, \dots, \Psi u_K)$$

and similarly also $\mathbf{F}_\rho\colon E \to Z$. In addition to the trivial choice $\Psi\colon E \to E$, $\Psi = \mathrm{Id}$ we consider discrete measurements in space-time, that is $Z \coloneqq \mathbb{R}^L$ and $(\Psi u)_l \approx u(y_l)$ for a finite number of sensor locations $y_1, \dots, y_L \in I \times \Omega$. The space $L^2(I; L^2(\Omega))$ does not allow point evaluations (and neither would $Y^{(k)}$). The formally correct way to proceed would be to set $(\Psi u)_l \coloneqq (\Psi * \varphi)(y_l)$ for some smooth function $\varphi\colon \mathbb{R}^{1+d} \to \mathbb{R}$ with small support. The corresponding adjoint reads

$$\Psi^*(w_1, \dots, w_L) = \sum_{l=1}^{L} w_l \varphi(\cdot - y_l).$$

In the case of a high number L of measurement locations, this approach induces a lot of computational overhead: For the evaluation of Ψ we would need to compute a lot of inner product with shifted (in space and time) versions of φ using numerical quadrature. Furthermore, for the adjoint we would have to interpolate these shifted functions onto the finite element space. The latter only has to be done once, but the storage requirements quickly exceed the size of all involved finite element matrices. Even for a few hundred sensor locations, we have observed that the measurement operator dominated the computation times. Much more convenient, although formally incorrect, is to simply evaluate the finite element function u at $y_1, \dots, y_L$. This is a much more local operation on the finite element space and can therefore be performed very quickly. The same holds for the adjoint, where one needs to project shifted versions of the δ-distribution onto the finite element space.

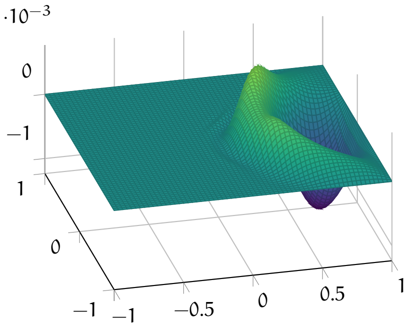
distribution of sensors We consider two possibilities how to distribute the sensor locations $y_1, \dots, y_L$ inside $I \times \Omega$. In both cases, we assume that the temporal resolution of the sensors is very high, and equal to the step size of our time-discretization, which yields 255 values at each measurement point in space. The first setting we would like to investigate are internal measurements, distributed on a uniform grid with distance 0.1 between nodes. Due to the Dirichlet boundary conditions we only take those nodes that lie inside of $(-1, 1)^2$, i.e.

$$\left\{ (-0.9 + 0.1i, -0.9 + 0.1j) \;\middle|\; i, j = 0, \dots, 18 \right\} \subset \Omega.$$

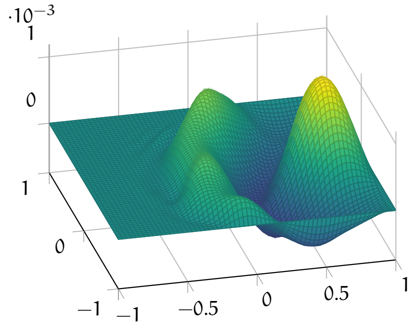
This defines 361 measurement locations inside of Ω, and a total of $L = 92055$ measurement values in $I \times \Omega$. Compared to full internal measurements this is a down-sampling to about 9% of previously available data points. Taking into account that $K = 4$, we end up with a nonlinear inverse problem with twice the number of unknowns in c or ρ as we have measurement values. For the second setting we only use the boundary nodes of this grid, i.e. 72 locations in space and $L = 18360$. Compared to full knowledge of u, this is only roughly 1.7% of the data.

For the ground-truths we again employ the function Λ_{LDot}, and use the $H^1(I; L^2(\Omega))$-norm for the reconstructions. Figure 8.18 contains graphical representations of the available data for the second right-hand side for the three different measurement setups. Even the full data, seen in Figure 8.18a, is much more challenging than it has been previously (cf. Figure 8.4), because each pulse barely reaches the opposing side of the domain during the time interval $I = [0, 2\pi]$. As expected, the grid-like distribution of sensors yields a massively down-sampled version of the full data, but it should still contain most of the information about what is happening inside of Ω. Obviously this is not the case anymore if we only consider the boundary nodes, as seen in Figure 8.18c.

reconstruction of the wave speed The reconstructions of c based on this data, tainted with 0.1% of noise, are the subject of Figure 8.19. Unsurprisingly, if we set Ψ to be the identity we obtain a good reconstruction, with a relative $H^1(I; L^2(\Omega))$-error of 37 %. If we use the down-sampled grid data, the error only increases slightly, to 42 %. However, here we see that the very fast computation of sensor values through point evaluations is not without cost: The δ-peaks at the sensor locations (introduced by Ψ^*) are clearly visible in the reconstruction. We have observed that a stronger norm like $H^1(I; H_0^1(\Omega))$ for the reconstruction process can be used to avoid these artifacts. This works because the cor-

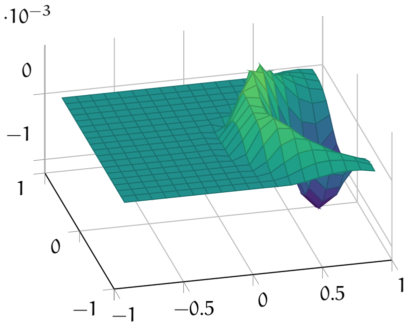


(a1) evaluated at $t = \pi$

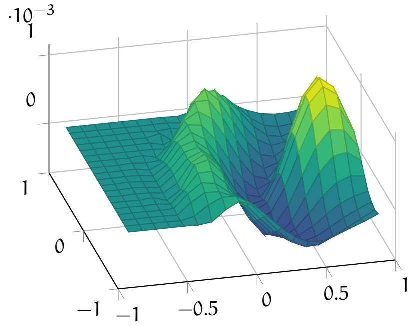


(a2) evaluated at $t = \frac{3}{2}\pi$

(a) $\Psi =$ identity

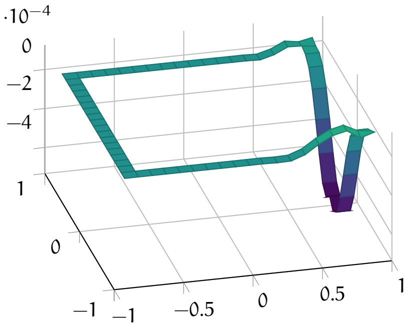


(b1) evaluated at $t = \pi$

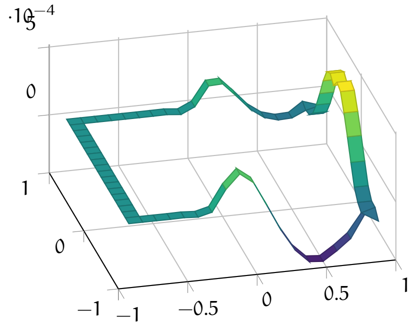


(b2) evaluated at $t = \frac{3}{2}\pi$

(b) $\Psi =$ internal measurements



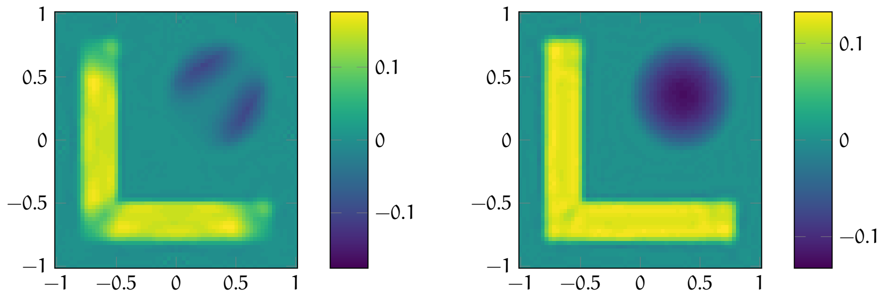
(c1) evaluated at $t = \pi$



(c2) evaluated at $t = \frac{3}{2}\pi$

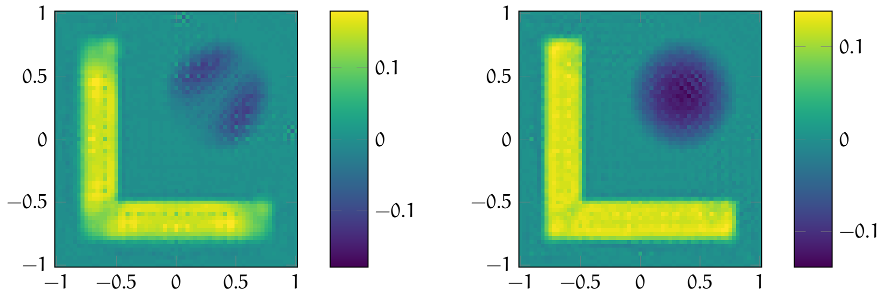
(c) $\Psi =$ boundary measurements

Figure 8.18: Data $[\mathbf{F}_c(0.15\Lambda_{\mathrm{LDot}})]_2$ for different measurement operators Ψ



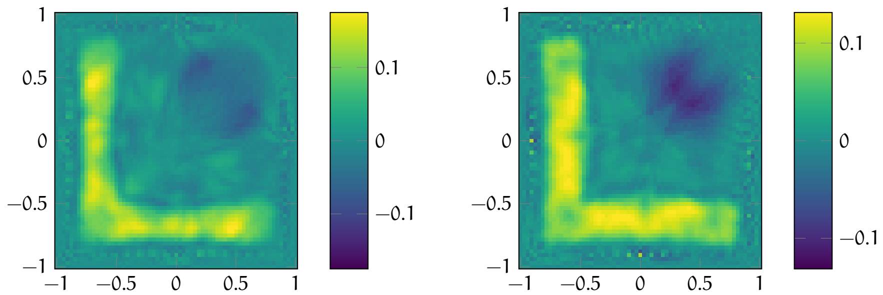
(a1) evaluated at $t = \frac{1}{2}\pi$ (a1) evaluated at $t = \frac{4}{3}\pi$

(a) $\Psi =$ identity



(b1) evaluated at $t = \frac{1}{2}\pi$ (b1) evaluated at $t = \frac{4}{3}\pi$

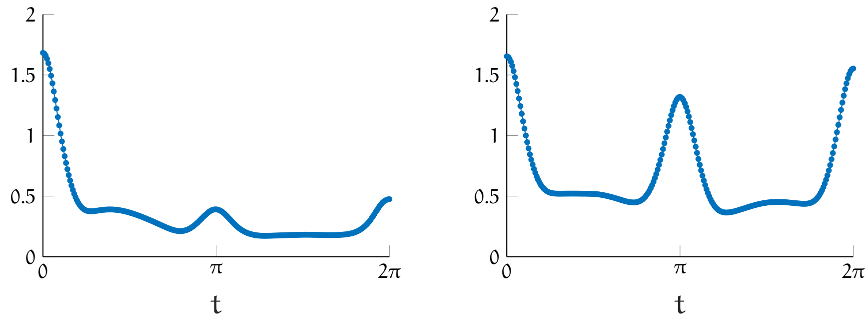
(b) $\Psi =$ internal measurements



(c1) evaluated at $t = \frac{1}{2}\pi$ (c1) evaluated at $t = \frac{4}{3}\pi$

(c) $\Psi =$ boundary measurements

Figure 8.19: Reconstruction of $c = 0.15\Lambda_{\mathrm{LDot}}$ using different measurement operators Ψ and 0.1% noise



(a) internal measurements (b) boundary measurements

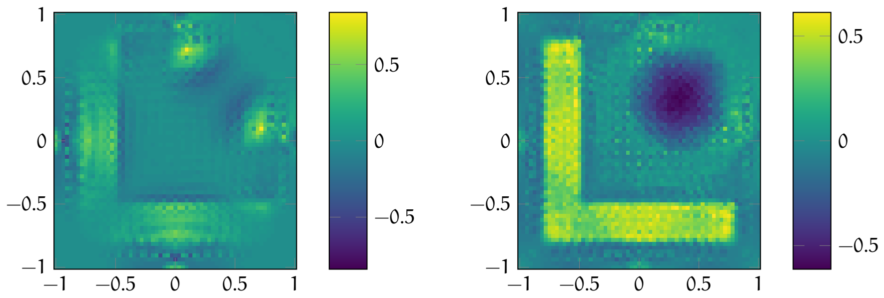
Figure 8.20: Relative $L^2(\Omega)$-reconstruction error over time for $c = 0.15\Lambda_{\text{LDot}}$ and 0.1% noise

responding transformation matrices perform a smoothing in the spatial variable (as discussed in Section 7.2.1).

The reconstruction from boundary data, seen in Figure 8.19c, achieves a relative error of 73 %. Although this error is very large, the reconstruction looks much better than expected. Both objects are clearly reconstructed. However, the discontinuities of the L-shape are not resolved as sharply as in the other reconstructions, and the ball-shaped object is perturbed. In addition, we still see the artifacts near the sensor locations, but here they are easily dealt with: They lie outside of the support of Λ_{LDot}, thus one could simply cut off this part of the reconstruction. For 1% of noise, the reconstructions look similar, and all of the errors increase by about 10%. All in all, these results for the wave speed are very promising with regards to actual applications, where coarse boundary data should be a common scenario.

ERROR OVER TIME Note that our setup heavily complicates the reconstruction of the parameter close to $t = 0$ and for boundary measurements also at $t = T$: Due to the local nature of our right-hand sides, the wave field u is not immediately active everywhere in the domain. Thus, changes in the parameters near the starting time have little to no effect on the discrepancy. Analogously, if we change the wave speed (or any other parameter) for $t \approx T$, then this change in the wave has not enough time to propagate to the sensors at the boundary. Indeed, this can be observed in the reconstructions: Figure 8.20 shows the relative $L^2(\Omega)$-error of each time frame of the reconstruction in case of grid- and boundary measurements. In both cases, the error starts out high. For the boundary measurements it severely increases again when we reach the end of the time interval I, whereas for a grid-like sensor distribution this growth is not as extreme.

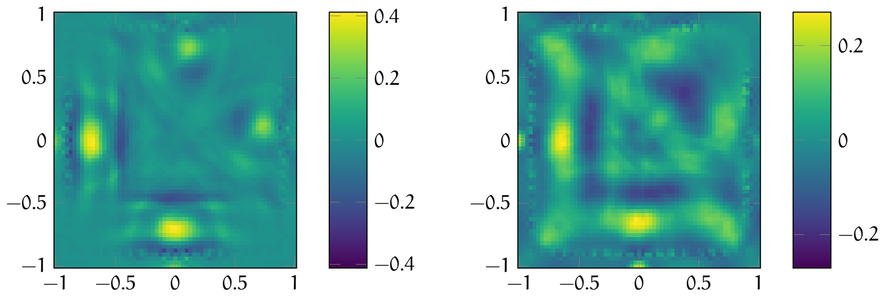
RECONSTRUCTION OF THE MASS DENSITY In almost all of the previous discussions, we found that ρ and c behave very similar. For $\Psi = \text{Id}$, this also turns out to be the case here, but the reconstructions from sensor data are of much worse quality for ρ than they are for c. The



(a) evaluated at $t = \frac{1}{2}\pi$

(b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.21: Reconstruction of $\rho = 0.75\Lambda_{\text{LDot}}$ from internal measurements and 0.1% noise



(a) evaluated at $t = \frac{1}{2}\pi$

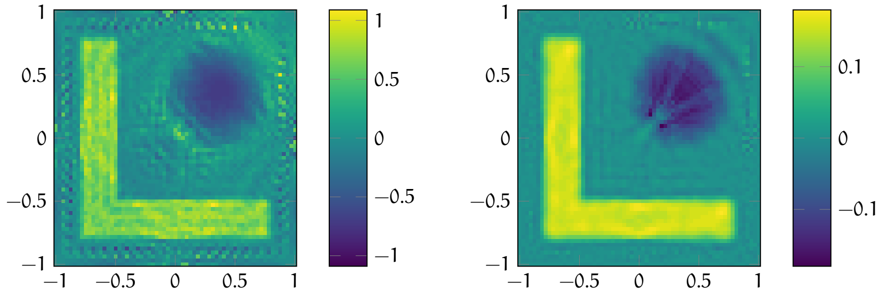
(b) evaluated at $t = \frac{4}{3}\pi$

Figure 8.22: Reconstruction of $\rho = 0.75\Lambda_{\text{LDot}}$ from boundary measurements and 0.1% noise

reconstruction from grid data is seen in Figure 8.21. The support of the L-shape is barely visible at $t = \pi/2$, and the ball is not detected at all. For $t = {}^4\!/_3\pi$ the image improves, but it is much more tainted by artifacts than the corresponding picture for c (Figure 8.19b). For boundary data the reconstruction totally breaks down, see Figure 8.22. Even for $t = {}^4\!/_3\pi$ we can hardly make out the supports of the objects. One possible explanation might be that even for full data, the initial discrepancy (based on the initial guess zero) in this scenario for F_ρ is much lower than F_c, they amount to 28 % and 53 %, i.e. the inversion of F_ρ will terminate sooner. For boundary data, the initial discrepancy for the ρ-problem is at only 18 %, whereas for c it was still at about 32 %. Together with the physical interpretation of ρ as the mass density and c as the wave speed, we are led to the hypothesis that changes in the wave speed produce more "global" effects (that are observable near the boundary), whereas changes in the mass density primarily cause local deviations.

8.4.1 *Static reconstructions*

We would like to take the low reconstruction quality for ρ for boundary data as an opportunity to show how the reconstruction quality can differ between the dynamic- and the static case. If we strip Λ_{LDot} of its time-dependence, i.e. use $\rho(x) = 0.75\Lambda_{\text{LDot}}(\pi/2, x)$ as the unknown, then the



(a) $\rho(x) = 0.75\Lambda_{\text{LDot}}(\pi/2, x)$ (b) $c(x) = 0.15\Lambda_{\text{LDot}}(\pi/2, x)$

Figure 8.23: Reconstructions of static parameters in $L^2(\Omega)$ in 2D with $\varepsilon = 10^{-3}$ from boundary measurements

reconstruction from boundary data is about as bad as already seen in Figure 8.22. This is not surprising, since the inversion algorithm does not know that the ground truth is constant in time.

Transforming our reconstruction algorithm for dynamic parameters to one for static parameters is not difficult, we mainly have to change the adjoint of ∂F, originally given by Theorem 5.5. If the parameters do not depend on time, then in (5.25) we can move the time integrals into the four dual products to obtain $L^\infty(\Omega) \times L^1(\Omega)$ pairings with the static unknowns. Thus, all we have to do is augment the original implementation by this time integral. We can even avoid having to change any data structures if we make the adjoint return $N + 1$ copies of this function. Apart from the initial guess (which we have set to zero), there is no other part of the algorithm that can introduce time-dependence into the reconstruction.

After these changes we obtain a much better image of ρ, which we have included as Figure 8.23a. This a-priori information about the ground truth turns the original $H^1(I; L^2(\Omega))$-error of 94 % into a $L^2(\Omega)$-error of only 32 %. We can still see the sensor positions, and inside the domain the image seems slightly noisy. The latter might indicate that using the discrepancy principle with $\tau = 2$ might have been a little too optimistic.

For c, the dynamic reconstruction of $c(x) = 0.15\Lambda_{\text{LDot}}(\pi/2, x)$, again based on boundary data, yields an image comparable to the already quite good reconstruction in Figure 8.19c. The corresponding static reconstruction is seen in Figure 8.23b and possesses a relative $L^2(\Omega)$-error of 17 %. Although both reconstructions in Figure 8.23 look satisfactory, even in the static case c seems to be easier to reconstruct from boundary measurements than ρ.

9 CONCLUSION & OUTLOOK

In this work we analyzed multiple dynamic inverse problems that involve second-order linear hyperbolic equations. In particular, we developed a general framework that allows to tackle such problems, and we demonstrated how it can be applied to various partial differential equations. Furthermore, we presented numerical examples for the acoustic wave equation, in which the task included the reconstruction of a time-dependent wave speed and mass density. These numerical examples demonstrated that if internal data about the wave field is available, then time-dependent parameters in a two- or three-dimensional setting can be reliably identified using reasonable computational resources. Surprisingly, we discovered that the determination of a dynamic wave speed is possible even from boundary measurements.

This work lays the foundation for a lot of possible further research, and we would like to close this thesis by outlining some of them.

First of all, one could employ adaptive meshes for the solution of the wave equation. Our PDE solver can already cope with time-dependent meshes. However, to transform the adjoint of ∂F to the correct inner product, we rely on the transformation matrices of Section 7.2, and in their current form they can only be implemented efficiently on a static mesh. This has to be remedied, for example by resorting to the analytical adjoint (which might impair the stability of the inversion). Naturally, the ability to adaptively refine and coarsen the mesh is useless without a criterion when to do so. Thus, one has to derive suitable error estimators for wave equations with time-dependent parameters, perhaps based on [BGR10; BR01; BS05], and decide how often the meshes should be changed. Note that in this case the finite element spaces for the unknown parameters and the wave fields should not be the same anymore, even more so when data for multiple right-hand sides is used. The space for the reconstruction can either be kept as is (linear elements on a uniform mesh), or it can be adaptively refined as well. Due to the high number of unknowns caused by the time-dependence of the parameters, this would be very beneficial to the storage requirements, in particular in three-dimensional scenarios. This could be achieved by building on the articles [Ban08; CKW16; KKV14a; KKV14b].

Related recent publications on dynamic inverse problems (e.g [Hah14; Hah17]) often include a more restricted model of the time-dependence. To be more precise, they assume that the searched for parameter is a static quantity that is perturbed by a time-dependent diffeomorphism. This effectively reduces the quest for a $d+1$-dimensional function to a d-dimensional function and a few scalars that control the deformation in time. This approach is already covered by our theoretical results, since it can be modeled by the value operator P. However, we have not analyzed the effect such a dimension-reduction has on the numerical reconstructions.

In this thesis we dealt with nonlinear inverse problems for linear partial differential equations. One could try to generalize this to nonlinear equations, preferably using a semilinear model as a starting point. The main difference will be that the forward operator then involves a fixed-point argument both for its theoretical analysis and the numerical implementation. Therefore, the analysis of such a problem can likely profit from our results for the forward operator of the corresponding linearized equation.

Lastly, we would like to mention that the classical scattering theory [CK13] usually treats inverse problems in which the data does not consist of only one scattered field. Instead, it uses the knowledge of which measurements an arbitrary excitation produces, often modeled by the operator that maps incident fields to far fields. In our numerical experiments we have already begun to approximate this setting by including a finite number of right-hand sides. As discussed in Section 5.2.1, moving to operator-valued data for the theoretical analysis would greatly increase the chances of proving injectivity of the forward operator, thereby guaranteeing uniqueness for the inverse problem.

We are confident that our theoretical analysis can provide a solid basis for these further considerations. In this spirit, we also made the source code relating to Chapters 7 and 8 publicly available at

`https://gitlab.informatik.uni-bremen.de/tgerken/wavepi`.

A

BOCHNER SPACES

Aside from the Introduction and the Outlook, all chapters of this thesis make heavy use of Bochner spaces. They extend the notions of the Lebesgue-integral and Sobolev spaces to functions that are additionally parameterized by a one-dimensional variable, which we often denote as the "time variable". They are very useful for the treatment of parabolic and hyperbolic partial differential equations, where one of the variables behaves differently from the others. Unfortunately, the results that we require for our analysis are scattered all over the literature, because most authors only state exactly those results they need for their purposes. We followed their example and have collected the definitions and assertions that are relevant for us in this appendix. For the proofs and more details we refer the reader to [Emm04; Eva10; Sch18; Yos95] and the appendix of [Zei90a].

Throughout this presentation we assume X to be a (real valued) Banach space and I to be the bounded interval $I := [0, T]$ with $T > 0$.

MEASURABILITY Similar to the Lebesgue integration theory, it is not possible to assign an integral to all $u \colon I \to X$.

Definition A.1. We make the following definitions for $u \colon I \to X$.

(i) u is *simple* if its image $u(I) \subset X$ is finite and all of its preimages are measurable, i.e. if it is of the form $u = \sum_{i=1}^{m} u_i \mathbb{1}_{E_i}$ with $m \in \mathbb{N}$, $u_1, \dots, u_m \in X$ and Lebesgue-measurable subsets $E_1, \dots, E_m$ of I.

(ii) u is *almost separably valued* if there exists $N \subset I$ of measure zero such that $u(I \setminus N) \subset X$ is separable.

(iii) u is *weakly measurable* if for all $f \in X^*$ the function $I \to \mathbb{R}$, $t \mapsto \langle f, u(t) \rangle$ is measurable (in the sense of Lebesgue).

(iv) u is *strongly measurable* (also *Bochner measurable*, or just *measurable*) if u can pointwise be approximated by simple functions, i.e. if there exist simple functions $u_k \colon I \to X$ $(k \in \mathbb{N})$ such that $u_k(t) \to u(t)$ in X as $k \to \infty$ for almost all $t \in I$.

Weak measurability is much easier to verify than strong measurability. For separable X these notions are equivalent, as the following theorem states. Of course this does not help with functions $I \to L^\infty(\Omega)$, like the unknown parameters in Chapters 5 and 6.

Theorem A.2 (Pettis). *$u \colon I \to X$ is strongly measurable if and only if it is weakly measurable and almost separably valued.*

Proof. See Section V.4 in YOSIDA [Yos95]. □

Definition A.3. For $p \in [1,\infty]$ let $\mathcal{L}^p(I;X)$ denote the set of all strongly measurable functions $u\colon I \to X$ for which

$$\|u\|_{L^p(I;X)} \coloneqq \begin{cases} \left(\int_0^T \|u(t)\|_X^p \, dt\right)^{1/p} & \text{if } p < \infty, \\ \operatorname*{ess\,sup}_{t\in I} \|u(t)\|_X & \text{if } p = \infty, \end{cases} \tag{A.1}$$

is finite. Further, we define $L^p(I;X) \coloneqq \mathcal{L}^p(I;X)/\sim$, with the equivalence relation that is defined by $f \sim g$ if $f = g$ almost everywhere.

Due to the properties of the Lebesgue integral, we immediately see $L^p(I;X) \subset L^q(I;X)$ if $1 \leqslant q \leqslant p \leqslant \infty$, and also note that $L^p(I;\mathbb{R}) = L^p(I)$. As is common for Lebesgue spaces, we also abuse notation for Bochner spaces by writing $u \in L^p(I;X)$ when we actually mean the equivalence class $[u] \in L^p(I;X)$ with a representative $u \in \mathcal{L}^p(I;X)$.

Even without an integral definition for functions $u\colon I \to X$, we can already deduce the following properties of the Bochner spaces.

Theorem A.4. *The following assertions hold for* $p \in [1,\infty]$.

(i) $L^p(I;X)$ *is a Banach space when equipped with the norm* (A.1)*. If* X *is a Hilbert space, then so is* $L^2(I;X)$*. A suitable inner product is for* $u, v \in L^2(I;X)$ *given by*

$$(u,v) \coloneqq \int_0^T (u(t), v(t))_X \, dt.$$

(ii) If X *is continuously embedded in another Banach space* Y*, then* $L^p(I;X)$ *is continuously embedded in* $L^q(I;Y)$ *for all* $q \in [1,p]$.

(iii) If X *is separable and* $p < \infty$*, then* $L^p(I;X)$ *is also separable.*

(iv) The space $C(I;X)$ *of continuous functions* $I \to X$ *(equipped with the supremum norm) is continuously embedded into* $L^p(I;X)$*. If* $p < \infty$ *then this embedding is dense.*

(v) Let $f \in L^q(I;X^*)$ *and* $u \in L^p(I;X)$ *with* $1/p + 1/q = 1$*. Then we have* $t \mapsto \langle f(t), u(t)\rangle \in L^1(I)$ *with*

$$\int_0^T |\langle f(t), u(t)\rangle| \, dt \leqslant \|f\|_{L^q(I;X^*)} \|u\|_{L^p(I;X)}.$$

(vi) If $p < \infty$ *and* X *is reflexive or has a separable dual space* X^**, then the dual space of* $L^p(I;X)$ *can be identified with* $L^q(I;X^*)$ $(1/p + 1/q = 1)$ *by*

$$\langle f, u\rangle \coloneqq \int_0^T \langle f(t), u(t)\rangle_{X^* \times X} \, dt, \quad f \in L^q(I;X^*), u \in L^p(I;X).$$

In particular, we conclude that $L^p(I;X)$ *is reflexive for* $p \in (1,\infty)$.

Proof. See Satz 7.1.23 in Emmrich [Emm04]. □

One can further show that for open $\Omega \subset \mathbb{R}^d$, the space $L^p(I;L^p(\Omega))$ is isomorphic to $L^p(I \times \Omega)$ if $p < \infty$. For $p = \infty$ only $L^\infty(I;L^\infty(\Omega)) \subset L^\infty(I \times \Omega)$ holds, because we cannot use Pettis' Theorem to show that the functions belonging to $L^\infty(I \times \Omega)$ provide strongly measurable functions $I \to L^\infty(\Omega)$.

INTEGRATION We can easily assign an integral to simple functions $u\colon I \to X$ by

$$\int_0^T u(t)\,dt := \sum_{x\in u(I)} x\,\lambda\big(u^{-1}(\{x\})\big),$$

where λ denotes the Lebesgue measure on I. If $u \in L^p(I;X)$ for some $p \in [1,\infty]$, then it can (by definition) be approximated by simple functions u_k, $k \in \mathbb{N}$. This suggests the definition

$$\int_0^T u(t)\,dt := \lim_{k\to\infty} \int_0^T u_k(t)\,dt, \tag{A.2}$$

which we call the *Bochner integral* of u and yields a value in X. If B is a measurable subset of I, then we define $\int_B u(t)\,dt$ to be the integral of $\mathbb{1}_B u$. Note that the latter function is also an element $L^p(I;X)$ because it satisfies $\|\mathbb{1}_B u\|_{L^p(I;X)} < \infty$ and can be approximated by the sequence $\mathbb{1}_B u_k$ that consists of simple functions. For $0 \leqslant s \leqslant t \leqslant T$ we further introduce the notation $\int_s^t u(t)\,dt := \int_{(s,t)} u(t)\,dt$, as well as $\int_t^s u(t)\,dt := -\int_s^t u(t)\,dt$.

The following theorem shows that the above integral definition makes sense, and contains some useful properties of the integral.

Theorem A.5. *Let $p \in [1,\infty)$ and $u \in L^p(I;X)$. Then the following assertions hold.*

(i) The limit in the integral definition (A.2) exists and is independent of the choice of the sequence $(u_k)_{k\in\mathbb{N}}$.

(ii) There is a sequence of simple functions $u_k\colon I \to X$ (with $k \in \mathbb{N}$) that converges pointwise almost everywhere to u and further satisfies

$$\int_0^T \|u(t) - u_k(t)\|_X^p\,dt \to 0 \quad \text{as } k \to \infty.$$

(iii) The inequality

$$\left\|\int_0^T u(t)\,dt\right\|_X \leqslant \int_0^T \|u(t)\|_X\,dt$$

holds, as does the equality

$$\left\langle f, \int_0^T u(t)\,dt\right\rangle = \int_0^T \langle f, u(t)\rangle\,dt \quad \text{for all } f \in X^*.$$

Proof. The assertion can be found as Theorem 10.4 in Schweizer [Sch18], alongside its proof. □

Moreover, we obtain the following version of the dominated convergence theorem, which is also taken from [Sch18].

Theorem A.6 (Lebesgue). *Let $p \in [1,\infty)$, $u\colon I \to X$, $u_k \in L^p(I;X)$, $g, g_k \in L^p(I)$ with $g_k \to g$ in $L^p(I)$ and $u_k(t) \to u(t)$ almost everywhere when $k \to \infty$. Suppose further that $\|u_k(t)\| \leqslant g_k(t)$ for almost all $t \in I$. Then $u \in L^p(I;X)$ and $u_k \to u$ with respect to the $L^p(I;X)$-norm.*

As one would expect, integration yields almost everywhere differentiable functions (in the classical sense).

Lemma A.7. *Let* $v \in L^1(I;X)$ *and* $t_0 \in I$*. The function* $u\colon I \to X$*, defined via*

$$u(t) := \int_{t_0}^{t} v(\tau)\,d\tau, \quad t \in I,$$

is continuous. Furthermore, it is differentiable in almost all $t \in I$*, with derivative* $u'(t) = v(t)$*. This also holds in all* t *in which* v *is continuous.*

Proof. See Satz 7.1.19 in Emmrich [Emm04]. □

WEAK DERIVATIVES The notion of integrals for X-valued functions also allows to define (weak) differentiability for them.

Definition A.8. Let $p \in [1, \infty]$ and $k \in \mathbb{N}_0$.

(i) Let $u \in L^p(I;X)$. By

$$u^{(k)}\colon C_c^\infty(I) \to X, \quad \varphi \mapsto (-1)^k \int_0^T u(t)\varphi^{(k)}(t)\,dt,$$

every such function possesses distributional derivatives of arbitrary high order k. We write $u^{(k)} \in L^p(I;X)$ and denote $u^{(k)}$ as the *weak derivative* of u of order k if $u^{(k)}$ is a regular distribution that can be represented by some $v \in L^p(I;X)$, that is if

$$(-1)^k \int_0^T u(t)\varphi^{(k)}(t)\,dt = \int_0^T v(t)\varphi(t)\,dt$$

holds for all $\varphi \in C_c^\infty(I)$. In this case we do not distinguish between $u^{(k)}$ and v.

(ii) We define the *Sobolev-Bochner space*

$$W^{k,p}(I;X) := \left\{ u \in L^p(I;X) \;\middle|\; u^{(i)} \in L^p(I;X) \text{ for all } 0 \leqslant i \leqslant k \right\},$$

which becomes a Banach space when equipped with the norm

$$\|u\|_{W^{k,p}(I;X)} := \begin{cases} \left(\displaystyle\sum_{i=0}^{k} \left\| u^{(i)} \right\|_{L^p(I;X)}^p \right)^{1/p} & \text{if } p < \infty, \\ \displaystyle\sup_{i=1,\dots,k} \left\| u^{(i)} \right\|_{L^p(I;X)} & \text{if } p = \infty. \end{cases}$$

If X is a Hilbert space, then so is $H^k(I;X) := W^{k,2}(I;X)$, for example with the inner product

$$(u,v)_{H^k(I;X)} := \sum_{i=0}^{k} \left(u^{(i)}, v^{(i)} \right)_{L^2(I;X)}, \qquad u, v \in H^k(I;X).$$

It is of little consequence how we combine the $L^p(I;X)$-norms of $u^{(i)}$ to form the norm of $u \in W^{k,p}(I;X)$; we could have employed any norm on $\mathbb{R}^k$ for this purpose. However, it seems to be the most "consistent" approach to also utilize the p-norm for this.

We can immediately see that weakly differentiable functions can be made classically differentiable through convolution with more regular functions.

Lemma A.9. *Let $(\eta_\varepsilon)_{\varepsilon>0} \subset C_c^\infty(I)$ a smoothing kernel, $p \in [1,\infty)$ and $u \in W^{1,p}(I;X)$. We extend u to $\mathbb{R}$ by 0 to define $u_\varepsilon\colon I \to X$ by*

$$u_\varepsilon(t) \coloneqq (u * \eta_\varepsilon)(t) = \int_{\mathbb{R}} u(s)\,\eta_\varepsilon(t-s)\,ds, \quad t \in I.$$

For all $\varepsilon > 0$ we obtain $u_\varepsilon \in C^\infty(I;X)$, and for $t \in (0,T)$ we can exchange smoothing and differentiation, i.e. $(u_\varepsilon)'(t) = (u')_\varepsilon(t)$. When $\varepsilon \to 0$, then $u_\varepsilon \to u$ with respect to the norm of $W^{1,p}(I;X)$. In particular, we observe that $C^\infty(I;X)$ is dense in $W^{1,p}(I;X)$.

Similar to one-dimensional Sobolev spaces, functions belonging to $W^{1,p}(I;X)$ are continuous after redefining them on a set of measure zero.

Theorem A.10. *Let $p \in [1,\infty]$. Each $u \in W^{1,p}(I;X)$ has a continuous representative $\tilde{u} \in C(I;X)$. It satisfies the estimate*

$$\max_{t\in I} \|\tilde{u}(t)\|_X \leqslant C\|u\|_{W^{1,p}(I;X)}$$

with $C > 0$ being independent of u. In particular, the embedding of $W^{1,p}(I;X)$ into $C(I;X)$ is well-defined and continuous. Furthermore, the fundamental theorem

$$\tilde{u}(t) - \tilde{u}(s) = \int_s^t u'(\tau)\,d\tau$$

is valid in all $t, s \in I$.

Proof. See Theorem 2 in Section 5.9 of Evans' book [Eva10]. □

Through $\tilde{u}$ we obtain a way to define point evaluations for $u \in W^{1,p}(I;X)$ while still maintaining continuous dependence on the whole equivalence class. Hence we choose to identify u with its continuous variant $\tilde{u}$ whenever possible.

If $u \in W^{2,p}(I;X)$, then u' is continuous as well and together with Lemma A.7 we are able to deduce that $u \in C^1(I;X)$. By induction we obtain for $k \in \mathbb{N}$ the continuous embedding $W^{k,p}(I;X) \hookrightarrow C^{k-1}(I;X)$.

CONNECTION TO GELFAND TRIPLES The structure of Gelfand triples $V \subset H \subset V^*$ (see Definition 2.1) allows to show continuity of functions $I \to V$ with respect to H, even if its derivative only maps to the worse space V^*. With such functions we are further able to integrate by parts.

Theorem A.11. *Let $V \subset H \subset V^*$ be a Gelfand triple and $p, q \in (1,\infty)$ with $1/p + 1/q = 1$.*

(i) *Every* $u \in L^p(I;V)$ *with* $u' \in L^q(I;V^*)$ *has a continuous representative* $\tilde{u} \in C(I;H)$. *It satisfies the estimate*

$$\max_{t\in I} \|\tilde{u}(t)\|_H \leqslant C\Big(\|u\|_{L^p(I;V)} + \|u'\|_{L^q(I;V^*)}\Big)$$

with a constant $C > 0$ *that does not depend on* u.

(ii) *For* $u, v \in L^p(I;V)$ *that have derivatives* $u', v' \in L^q(I;V^*)$, *the integration by parts formula*

$$(u(t), v(t)) - (u(s), v(s)) = \int_s^t \big\langle u'(\tau), v(\tau)\big\rangle + \big\langle v'(\tau), u(\tau)\big\rangle \, d\tau$$

holds for all $0 \leqslant s \leqslant t \leqslant T$. *Note that the evaluations of* u *and* v *on the left-hand side are well-defined in* H *because of assertion (i).*

Proof. Can be found as Proposition 23.23 in Volume 2A of Zeidler's monograph [Zei90b]. □

If X is a separable Hilbert space, then the integration by parts formula in particular holds for the Gelfand triple "$X \subset X \subset X$" and $p = q = 2$, i.e. for functions $u, v \in H^1(I;X)$. We also recover the continuity of the embedding $H^1(I;X) \hookrightarrow C(I;X)$ (cf. Theorem A.10) as a special case of assertion (i).

B NUMERICAL CONVERGENCE RATES

Tables B.1 to B.4 contain supplementary material of the analysis done in Section 8.3. Tables B.1 and B.2 contain rates for each forward operator in the two-dimensional setting. The former deals with Λ_{LDot} and the latter with Λ_{Ring} The tables are to be read as follows: For each combination of a forward operator F_c, F_ν, F_ρ, F_q and possible reconstruction norm, the reconstruction errors behave like $\mathcal{O}(\varepsilon^\gamma)$ with γ as given in the corresponding table cell. In addition to the behavior of the errors, we also included the rate in which the amount of partial differential equations that have to be solved increases as ε tends to zero. Note that due to higher computing times in 3D, Tables B.3 and B.4 only contain results where the $H^1(I;L^2(\Omega))$-norm is used for the domain of definition. The values of ε we used are the same as in Section 8.3. The lowest noise levels are 10^{-4} in 2D and 10^{-3} in 3D. In the case that some reconstructions did not converge (either due to divergence or a timeout after multiple days of wall time), we have indicated the lowest noise level for which a solution could be obtained.

Type	Norm	Convergence Rates $L^2(I;L^2)$	$H^1(I;L^2)$	$H^2(I;L^2)$	PDEs	Remarks
c	$L^2(I;L^2)$	0.15	−0.70	−4.53	−3.40	$\varepsilon \geqslant 1.0 \cdot 10^{-1}$
	$H^1(I;L^2)$	0.31	0.27	0.06	−0.78	
	$H^2(I;L^2)$	0.32	0.29	0.21	−0.78	$\varepsilon \geqslant 1.0 \cdot 10^{-3}$
ν	$L^2(I;L^2)$	0.13	−0.10	−0.38	−0.84	
	$H^1(I;L^2)$	0.28	0.22	0.06	−0.78	
	$H^2(I;L^2)$	0.29	0.25	0.17	−0.78	
ρ	$L^2(I;L^2)$	0.02	−0.44	−1.33	−1.67	$\varepsilon \geqslant 1.0 \cdot 10^{-2}$
	$H^1(I;L^2)$	0.23	0.16	−0.02	−0.78	
	$H^2(I;L^2)$	0.34	0.27	0.16	−0.86	
q	$L^2(I;L^2)$	0.14	−0.08	−0.36	−0.80	
	$H^1(I;L^2)$	0.26	0.21	0.07	−0.75	
	$H^2(I;L^2)$	0.27	0.24	0.16	−0.74	

Table B.1: Rates for Λ_{LDot} in 2D

Type	Norm	Convergence Rates $L^2(I;L^2)$	$H^1(I;L^2)$	$H^2(I;L^2)$	PDEs	Remarks
c	$L^2(I;L^2)$	0.08	−0.50	−0.57	−3.66	$\varepsilon \geqslant 4.6 \cdot 10^{-2}$
	$H^1(I;L^2)$	0.31	0.05	0.00	−1.34	$\varepsilon \geqslant 1.0 \cdot 10^{-2}$
	$H^2(I;L^2)$	0.26	0.03	0.00	−1.75	$\varepsilon \geqslant 2.2 \cdot 10^{-2}$
v	$L^2(I;L^2)$	0.13	0.00	0.00	−0.93	
	$H^1(I;L^2)$	0.22	0.03	0.00	−1.10	
	$H^2(I;L^2)$	0.19	0.03	0.00	−1.42	
ρ	$L^2(I;L^2)$	0.03	−0.18	−0.17	−1.46	$\varepsilon \geqslant 4.6 \cdot 10^{-3}$
	$H^1(I;L^2)$	0.19	0.03	0.00	−0.90	$\varepsilon \geqslant 1.0 \cdot 10^{-3}$
	$H^2(I;L^2)$	0.17	0.02	0.00	−0.98	$\varepsilon \geqslant 4.6 \cdot 10^{-3}$
q	$L^2(I;L^2)$	0.13	0.00	−0.00	−0.85	
	$H^1(I;L^2)$	0.22	0.03	0.00	−1.03	
	$H^2(I;L^2)$	0.21	0.03	0.00	−1.37	

Table B.2: Rates for Λ_{Ring} in 2D

Type	Convergence Rates $L^2(I;L^2)$	$H^1(I;L^2)$	$H^2(I;L^2)$	PDEs	Remarks
c	0.31	0.23	0.06	−0.72	$\varepsilon \geqslant 1.0 \cdot 10^{-2}$
v	0.27	0.20	0.05	−0.89	
ρ	0.23	0.16	0.02	−0.90	
q	0.28	0.22	0.07	−0.89	

Table B.3: Rates for Λ_{LDot} using $H^1(I;L^2(\Omega))$ in 3D

Type	Convergence Rates $L^2(I;L^2)$	$H^1(I;L^2)$	$H^2(I;L^2)$	PDEs	Remarks
c	0.30	0.04	0.00	−1.73	$\varepsilon \geqslant 2.2 \cdot 10^{-2}$
v	0.20	0.02	0.00	−0.98	
ρ	0.14	0.02	0.00	−1.03	$\varepsilon \geqslant 2.2 \cdot 10^{-3}$
q	0.21	0.03	0.00	−1.00	

Table B.4: Rates for Λ_{Ring} using $H^1(I;L^2(\Omega))$ in 3D

BIBLIOGRAPHY

[AF03] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. 2nd ed. Pure and Applied Mathematics. Academic Press, 2003.

[Aïc15] I. B. Aïcha. "Stability Estimate for a Hyperbolic Inverse Problem with Time-Dependent Coefficient." In: *Inverse Problems* 31.12 (2015), p. 125010.

[Arn+19] D. Arndt et al. "The `deal.II` Library, Version 9.1." In: *Journal of Numerical Mathematics* (2019).

[Aub63] J.-P. Aubin. "Un théorème de compacité." French. In: *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 256 (1963), pp. 5042–5044.

[BGR10] W. Bangerth, M. Geiger, and R. Rannacher. "Adaptive Galerkin Finite Element Methods for the Wave Equation." In: *Computational Methods in Applied Mathematics* 10.1 (2010), pp. 3–48.

[BHK07] W. Bangerth, R. Hartmann, and G. Kanschat. "`deal.II` — A General-Purpose Object-Oriented Finite Element Library." In: *ACM Transactions on Mathematical Software* 33.4 (2007).

[BR99] W. Bangerth and R. Rannacher. "Finite Element Approximation of the Acoustic Wave Equation: Error Control and Mesh Adaptation." In: *East-West Journal of Numerical Mathematics* 7.4 (1999), pp. 263–282.

[Ban08] W. Bangerth. "A Framework for the Adaptive Finite Element Solution of Large-Scale Inverse Problems." In: *SIAM Journal on Scientific Computing* 30.6 (2008), pp. 2965–2989.

[BR01] W. Bangerth and R. Rannacher. "Adaptive Finite Element Techniques for the Acoustic Wave Equation." In: *Journal of Computational Acoustics* 09.02 (2001), pp. 575–591.

[BR03] W. Bangerth and R. Rannacher. *Adaptive Finite Element Methods for Differential Equations*. Lectures in Mathematics ETH Zürich. Basel: Birkhäuser, 2003.

[BB19] M. Bellassoued and I. Ben Aïcha. "An Inverse Problem of Finding Two Time-Dependent Coefficients in Second Order Hyperbolic Equations from Dirichlet to Neumann Map." In: *Journal of Mathematical Analysis and Applications* 475.2 (2019), pp. 1658–1684.

[BS05] C. Bernardi and E. Süli. "Time and Space Adaptivity for the Second-Order Wave Equation." In: *Mathematical Models and Methods in Applied Sciences* 15.02 (2005), pp. 199–225.

[BSS13] K. D. Blazek, C. Stolk, and W. W. Symes. "A Mathematical Framework for Inverse Wave Problems in Heterogeneous Media." In: *Inverse Problems* 29.6 (2013), p. 065001.

[BS08] S. Brenner and R. Scott. *The Mathematical Theory of Finite Element Methods*. 3rd ed. Texts in Applied Mathematics 15. New York: Springer, 2008.

[Bré11] H. Brézis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. New York: Springer, 2011.

[BW13] C. Büskens and D. Wassel. "The ESA NLP Solver WORHP." In: *Modeling and Optimization in Space Engineering*. Ed. by G. Fasano and J. D. Pintér. Springer Optimization and Its Applications 73. New York: Springer, 2013, pp. 85–110.

[CP11] A. Chambolle and T. Pock. "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging." In: *Journal of Mathematical Imaging and Vision* 40.1 (2011), pp. 120–145.

[Cia88] P. G. Ciarlet. *Mathematical Elasticity Volume I: Three-Dimensional Elasticity*. Studies in Mathematics and Its Applications 20. Amsterdam: North-Holland, 1988.

[CL05] R. Cipolatti and I. F. Lopez. "Determination of Coefficients for a Dissipative Wave Equation via Boundary Measurements." In: *Journal of Mathematical Analysis and Applications* 306.1 (2005), pp. 317–329.

[CKW16] C. Clason, B. Kaltenbacher, and D. Wachsmuth. "Functional Error Estimators for the Adaptive Discretization of Inverse Problems." In: *Inverse Problems* 32.10 (2016), p. 104004.

[CL55] E. A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. New York: McGraw-Hill, 1955.

[CK13] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. 3rd ed. Applied Mathematical Sciences 93. New York: Springer, 2013.

[DL00] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology Volume V: Evolution Problems I*. Berlin, Heidelberg: Springer, 2000.

[Dav04] T. A. Davis. "Algorithm 832: UMFPACK V4.3 – an Unsymmetric-Pattern Multifrontal Method." In: *ACM Transactions on Mathematical Software* 30.2 (2004), pp. 196–199.

[EGK17] C. Eck, H. Garcke, and P. Knabner. *Mathematische Modellierung*. German. 3rd ed. Springer-Lehrbuch. Springer Spektrum, 2017.

[EH18] T. Elfving and P. Hansen. "Unmatched Projector/Backprojector Pairs: Perturbation and Convergence Analysis." In: *SIAM Journal on Scientific Computing* 40.1 (2018), A573–A591.

[Emm04] E. Emmrich. *Gewöhnliche und Operator-Differentialgleichungen*. German. Wiesbaden: Vieweg+Teubner, 2004.

[EHN00] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications 375. Springer Netherlands, 2000.

[Esk07] G. Eskin. "Inverse Hyperbolic Problems with Time-Dependent Coefficients." In: *Communications in Partial Differential Equations* 32.11 (2007), pp. 1737–1758.

[Eva10] L. C. Evans. *Partial Differential Equations*. 2nd ed. Graduate Studies in Mathematics 19. Providence, R.I.: American Mathematical Society, 2010.

[Fey15] R. P. Feynman. *The Feynman Lectures on Physics II: Mainly Electromagnetism and Matter*. New York: Basic Books, 2015.

[For88] B. Fornberg. "Generation of Finite Difference Formulas on Arbitrarily Spaced Grids." In: *Mathematics of Computation* 51.184 (1988), pp. 699–706.

[Ger16] T. Gerken. "Identifikation eines zeitabhängigen Parameters einer Wellengleichung." German. Master Thesis. University of Bremen, 2016.

[Ger19] T. Gerken. "Dynamic Inverse Wave Problems – Part II: Operator Identification and Applications." In: *Inverse Problems* (2019).

[GG19] T. Gerken and S. Grützner. "Dynamic Inverse Wave Problems – Part I: Regularity for the Direct Problem." In: *Inverse Problems* (2019).

[GL17] T. Gerken and A. Lechleiter. "Reconstruction of a Time-Dependent Potential from Wave Measurements." In: *Inverse Problems* 33.9 (2017), p. 094001.

[GR05] C. Grossmann and H.-G. Roos. *Numerische Behandlung partieller Differentialgleichungen*. German. 3rd ed. Teubner Studienbücher Mathematik. Vieweg+Teubner, 2005.

[GM18] S. Grützner and A. Muntean. "Brief Introduction to Damage Mechanics and Its Relation to Deformations." In: *Mathematical Analysis of Continuum Mechanics and Industrial Applications II: Proceedings of the International Conference CoMFoS16*. Ed. by P. van Meurs, M. Kimura, and H. Notsu. Mathematics for Industry 30. Singapore: Springer Nature, 2018.

[Hah13] B. N. Hahn. "Reconstruction of Dynamic Objects with Affine Deformations in Computerized Tomography." In: *Journal of Inverse and Ill-posed Problems* 22.3 (2013), pp. 323–339.

[Hah14] B. N. Hahn. "Efficient Algorithms for Linear Dynamic Inverse Problems with Known Motion." In: *Inverse Problems* 30.3 (2014), p. 035008.

[Hah17] B. N. Hahn. "A Motion Artefact Study and Locally Deforming Objects in Computerized Tomography." In: *Inverse Problems* 33.11 (2017), p. 114001.

[HS01] M. Hanke and O. Scherzer. "Inverse Problems Light: Numerical Differentiation." In: *The American Mathematical Monthly* 108.6 (2001), pp. 512–521.

[Hay+16] A. G. Hayrapetyan, J. B. Götte, K. K. Grigoryan, S. Fritzsche, and R. G. Petrosyan. "Electromagnetic Wave Propagation in Spatially Homogeneous yet Smoothly Time-Varying Dielectric Media." In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 178 (2016), pp. 158–166.

[HS94] B. Hofmann and O. Scherzer. "Factors Influencing the Ill-Posedness of Nonlinear Problems." In: *Inverse Problems* 10.6 (1994), p. 1277.

[HK18] G. Hu and Y. Kian. "Determination of Singular Time-Dependent Coefficients for Wave Equations from Full and Partial Data." In: *Inverse Problems and Imaging* 12.3 (2018), p. 745.

[Ika00] M. Ikawa. *Hyperbolic Partial Differential Equations and Wave Phenomena*. Translations of Mathematical Monographs 189. Providence, R.I.: American Mathematical Society, 2000.

[IS92] V. Isakov and Z. Sun. "Stability Estimates for Hyperbolic Inverse Problems with Local Boundary Data." In: *Inverse Problems* 8.2 (1992), p. 193.

[Jen+11] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt. *Computational Ocean Acoustics*. 2nd ed. Modern Acoustics and Signal Processing. New York: Springer, 2011.

[KKV14a] B. Kaltenbacher, A. Kirchner, and S. Veljović. "Goal Oriented Adaptivity in the IRGNM for Parameter Identification in PDEs: I. Reduced Formulation." In: *Inverse Problems* 30.4 (2014), p. 045001.

[KKV14b] B. Kaltenbacher, A. Kirchner, and B. Vexler. "Goal Oriented Adaptivity in the IRGNM for Parameter Identification in PDEs: II. All-at-Once Formulations." In: *Inverse Problems* 30.4 (2014), p. 045002.

[Kat88] T. Katō. *Abstract Differential Equations and Nonlinear Mixed Problems*. Publications of the Scuola Normale Superiore. Edizioni della Normale, 1988.

[Kia17] Y. Kian. "Unique Determination of a Time-Dependent Potential for Wave Equations from Partial Data." In: *Annales de l'Institut Henri Poincaré, Non Linear Analysis* 34.4 (2017), pp. 973–990.

[KO17] Y. Kian and L. Oksanen. "Recovery of Time-Dependent Coefficient on Riemannian Manifold for Hyperbolic Equations." In: *International Mathematics Research Notices* (2017).

[Kir11] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. 2nd ed. Applied Mathematical Sciences 120. New York: Springer, 2011.

[KH15] A. Kirsch and F. Hettlich. *The Mathematical Theory of Time-Harmonic Maxwell's Equations: Expansion-, Integral-, and Variational Methods*. Applied Mathematical Sciences 190. Springer International Publishing, 2015.

[KR14a] A. Kirsch and A. Rieder. "On the Linearization of Operators Related to the Full Waveform Inversion in Seismology." In: *Mathematical Methods in the Applied Sciences* 37.18 (2014), pp. 2995–3007.

[KR14b] A. Kirsch and A. Rieder. "Seismic Tomography Is Locally Ill-Posed." In: *Inverse Problems* 30.12 (2014), p. 125001.

[KR16] A. Kirsch and A. Rieder. "Inverse Problems for Abstract Evolution Equations with Applications in Electrodynamics and Elasticity." In: *Inverse Problems* 32.8 (2016), p. 085001.

[LR06] A. Lechleiter and A. Rieder. "Newton Regularizations for Impedance Tomography: A Numerical Study." In: *Inverse Problems* 22.6 (2006), p. 1967.

[LS17] A. Lechleiter and J. W. Schlasche. "Identifying Lamé Parameters from Time-Dependent Elastic Wave Measurements." In: *Inverse Problems in Science and Engineering* 25.1 (2017), pp. 2–26.

[Les91] R. A. Leslie. "How Not to Repeatedly Differentiate a Reciprocal." In: *The American Mathematical Monthly* 98.8 (1991), pp. 732–735.

[Lio61] J. L. Lions. *Equations Differentielles Operationnelles: Et Problèmes aux Limites*. French. Grundlehren der mathematischen Wissenschaften 111. Berlin, Heidelberg: Springer, 1961.

[LM72] J. L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications: Volume I*. Grundlehren Der Mathematischen Wissenschaften 181. Berlin, Heidelberg: Springer, 1972.

[MRL14] F. Margotti, A. Rieder, and A. Leitão. "A Kaczmarz Version of the REGINN-Landweber Iteration for Ill-Posed Problems in Banach Spaces." In: *SIAM Journal on Numerical Analysis* 52.3 (2014), pp. 1439–1465.

[Mar18] F. Margotti. "Inexact Newton Regularization Combined with Gradient Methods in Banach Spaces." In: *Inverse Problems* 34.7 (2018), p. 075007.

[McL00] W. C. H. McLean. *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge: Cambridge University Press, 2000.

[Mon03] P. Monk. *Finite Element Methods for Maxwell's Equations*. Oxford, New York: Oxford University Press, 2003.

[ORe97] D. O'Regan. *Existence Theory for Nonlinear Ordinary Differential Equations*. Mathematics and Its Applications 398. Dordrecht: Springer Netherlands, 1997.

[Paz83] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences 44. New York: Springer, 1983.

[RS91] A. G. Ramm and J. Sjöstrand. "An Inverse Problem of the Wave Equation." In: *Mathematische Zeitschrift* 206.1 (1991), pp. 119–130.

[RR91] A. G. Ramm and Rakesh. "Property C and an Inverse Problem for a Hyperbolic Equation." In: *Journal of Mathematical Analysis and Applications* 156.1 (1991), pp. 209–219.

[Rie99] A. Rieder. "On the Regularization of Nonlinear Ill-Posed Problems via Inexact Newton Iterations." In: *Inverse Problems* 15.1 (1999), p. 309.

[Rie01] A. Rieder. "On Convergence Rates of Inexact Newton Regularizations." In: *Numerische Mathematik* 88.2 (2001), pp. 347–365.

[Rie03] A. Rieder. *Keine Probleme mit Inversen Problemen: Eine Einführung in ihre stabile Lösung*. German. Wiesbaden: Vieweg+Teubner, 2003.

[Rie05] A. Rieder. "Inexact Newton Regularization Using Conjugate Gradients as Inner Iteration." In: *SIAM Journal on Numerical Analysis* 43.2 (2005), pp. 604–622.

[SS05] A. Schmidt and K. G. Siebert. *Design of Adaptive Finite Element Software: The Finite Element Toolbox ALBERTA*. Lecture Notes in Computational Science and Engineering 42. Berlin, Heidelberg: Springer, 2005.

[SLS06] F. Schöpfer, A. K. Louis, and T. Schuster. "Nonlinear Iterative Methods for Linear Ill-Posed Problems in Banach Spaces." In: *Inverse Problems* 22.1 (2006), p. 311.

[SHB18] T. Schuster, B. N. Hahn, and M. Burger. "Dynamic Inverse Problems: Modelling — Regularization — Numerics." In: *Inverse Problems* 34.4 (2018), p. 040301.

[Sch+12] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization Methods in Banach Spaces*. Radon Series on Computational and Applied Mathematics 10. Berlin, Boston: De Gruyter, 2012.

[Sch18] B. Schweizer. *Partielle Differentialgleichungen: Eine anwendungsorientierte Einführung*. German. 2nd ed. Masterclass. Springer Spektrum, 2018.

[Sho97] R. E. Showalter. *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*. Mathematical Surveys and Monographs 49. Providence, R.I.: American Mathematical Society, 1997.

[Sim86] J. Simon. "Compact Sets in the Space $L^p(0,T;B)$." In: *Annali di Matematica Pura ed Applicata* 146.1 (1986), pp. 65–96.

[Ste89] P. D. Stefanov. "Uniqueness of the Multi-Dimensional Inverse Scattering Problem for Time Dependent Potentials." In: *Mathematische Zeitschrift* 201.4 (1989), pp. 541–559.

[SB05] J. Stoer and R. Bulirsch. *Numerische Mathematik 2: Eine Einführung - unter Berücksichtigung von Vorlesungen von F.L. Bauer*. German. 5th ed. Springer-Lehrbuch. Berlin, Heidelberg: Springer, 2005.

[Sto01] C. Stolk. "On the Modeling and Inversion of Seismic Data." PhD Thesis. Utrecht University, 2001.

[Sym09] W. W. Symes. "The Seismic Reflection Inverse Problem." In: *Inverse Problems* 25.12 (2009), p. 123008.

[TM05] R. Temam and A. Miranville. *Mathematical Modeling in Continuum Mechanics*. 2nd ed. Cambridge, New York: Cambridge University Press, 2005.

[Wal18] A. Wald. "A Fast Subspace Optimization Method for Nonlinear Inverse Problems in Banach Spaces with an Application in Parameter Identification." In: *Inverse Problems* 34.8 (2018), p. 085008.

[Wat14] A. Waters. "Stable Determination of X-Ray Transforms of Time Dependent Potentials from Partial Boundary Data." In: *Communications in Partial Differential Equations* 39.12 (2014), pp. 2169–2197.

[Win16] R. Winkler. "A Model-Aware Inexact Newton Scheme for Electrical Impedance Tomography." PhD Thesis. Karlsruhe Institute of Technology, 2016.

[Wlo87] J. Wloka. *Partial Differential Equations*. Cambridge, London: Cambridge University Press, 1987.

[Yos95] K. Yosida. *Functional Analysis*. 6th ed. Classics in Mathematics. Berlin, Heidelberg: Springer, 1995.

[Zei88] E. Zeidler. *Nonlinear Functional Analysis and Its Applications Volume IV: Applications to Mathematical Physics*. New York: Springer, 1988.

[Zei90a] E. Zeidler. *Nonlinear Functional Analysis and Its Applications Volume II/A: Linear Monotone Operators*. New York: Springer, 1990.

[Zei90b] E. Zeidler. *Nonlinear Functional Analysis and Its Applications Volume II/B: Nonlinear Monotone Operators*. New York: Springer, 1990.